HUMAN GENETIC ANCESTRY, HEALTH, AND ADAPTATION IN LATIN AMERICA

A Dissertation
Presented to
The Academic Faculty

by

Emily T. Norris

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in Bioinformatics in the
School of Biological Sciences

Georgia Institute of Technology
December 2019

COPYRIGHT © 2019 BY EMILY T. NORRIS

HUMAN GENETIC ANCESTRY, HEALTH, AND ADAPTATION IN LATIN AMERICA

Approved by:

Dr. I. King Jordan, Advisor
School of Biological Sciences
Georgia Institute of Technology

Dr. Joseph Lachance
School of Biological Sciences
Georgia Institute of Technology

Dr. Gregory Gibson
School of Biological Sciences
Georgia Institute of Technology

Dr. John Lindo
Department of Anthropology
Emory University

Dr. Soojin Yi
School of Biological Sciences
Georgia Institute of Technology

Date Approved: October 23, 2019

To my family and friends

ACKNOWLEDGEMENTS

I am extremely grateful for my advisor Dr. I. King Jordan for his continuous support and guidance throughout my time as a student. His passion and commitment to science is inspiring and has helped me to be a better scientist. He is a wonderful mentor and throughout this program has guided me to be a better mentor to new students and a stronger person. I would not have been nearly as successful over the past few years without his encouragement.

I am also grateful for the guidance I have received from my committee members – Dr. Greg Gibson, Dr. Soojin Yi, Dr. Joe Lachance, and Dr. John Lindo. The four of you have provided me with helpful discussions and insights on my work and have encouraged me throughout this program to become a better scientist.

I also thank my friends and colleagues Michael Astwood and Feda Masseoud for supporting me during my PhD and for their constant belief in me. It has helped me to believe more in myself and become a stronger person. I must also thank my friends and colleagues in ABiL and the Jordan lab for their constant motivation and assistance in helping me achieve my goals.

I would like to thank Lisa Redding for her assistance with every question I had relating to the program and program-related requirements. I knew that if I ever had a concern Lisa would either know the answer or find the answer for me right away. Lisa has made the administrative side of this program simple for me and I cannot thank her enough for that and her consistent encouragement. I also thank Troy Hilley for all of his IT support

throughout the program. Without his assistance I would not have been able finish in the time that I did.

I am extremely appreciative of Kiera Berger and Michelle Kim for being some of the greatest friends and support system I could ask for throughout my PhD program. Meeting you both during orientation has provided me with wonderful friends with whom to celebrate my successes and commiserate over my failures. Along the same lines, I must thank Meagan Loew for being a great friend for the last 12+ years and for always having the time to talk even when life has taken us to different places.

Finally, I would not have been able to do this without the love and support of Lava Rishishwar, my parents John and Ellen Norris, and my sister Katy Norris. The four of you have always supported my decisions and encouraged me to continue with my education and are the reason I have worked hard to finish my degree.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

1KGP	1000 Genomes Project
AMH	Anatomically Modern Humans
BSP	Bantu-speaking Population
CLM	Colombian in Medellin, Colombia
eQTL	Expression Quantitative Trait Loci
FDR	False Discovery Rate
GIANT	Genetic Investigation of ANthropometric Traits
GSEA	Gene Set Enrichment Analysis
HGDP	Human Genome Diversity Project
HLA	Human Leukocyte Antigen
HW	Hard-Weinberg
iHS	Integrated Haplotype Score
kya	Thousand Years Ago
LA	Latin America
LD	Linkage Disequilibrium
MHC	Major Histocompatibility Complex
MsigDB	Molecular Signatures Database
MXL	Mexican Ancestry in Los Angeles, California
PEL	Peruvian in Lima, Peru
PUR	Puerto Rican in Puerto Rico
SNP	Single Nucleotide Polymorphism
ya	Years Ago

SUMMARY

This PhD thesis explores the implications of genetic admixture for human health, fitness, and population structure in Latin America. The motivation behind this work is rooted in the story of human evolution, with our origins in Africa followed by migration around the world. Admixture has been a constant and ubiquitous feature of human evolution throughout this time, but its impact has yet to be systematically explored. Latin America provides an ideal setting to explore the implications of admixture given the formation of modern populations via admixture among distinct African, European, and Native American population groups. My research into this subject is powered by comparative analysis of thousands of whole genome sequences integrated with a variety of functional genomic and clinical genetic data sources.

Anatomically modern humans evolved in Africa ~250,000 years ago and began to migrate out of Africa ~75,000 years ago, forming new, isolated populations in geographically distinct areas of the world. At that time, geographic isolation of human populations entailed reproductive isolation, and as a consequence, remote populations accumulated genetic differences. Most of the genetic divergence among isolated human populations is likely to have accumulated due to genetic drift, i.e. random changes in allele frequencies that have no bearing on fitness, but natural selection also played an important role in driving population differences. This is because humans encountered vastly different environmental conditions, and accordingly distinct selection pressures, on their long march out of Africa and around the world. Natural selection enabled human populations to adapt to their local environments by increasing the frequency of beneficial genetic variants.

Human population-specific adaptations include alleles that confer resistance to malaria in West Africa, lactose tolerance in Europe and East Africa, and the ability to thrive in the low oxygen, high altitude environments of Tibet in Asia and the Andes in South America.

Genetic admixture is the process that occurs when populations that were previously reproductively isolated, and consequently genetically diverged, come back together and exchange genes. Recent studies of modern and ancient genomes have underscored the frequency with which admixture has occurred during human evolution. Indeed, human evolution has been characterized by numerous iterations of physical isolation and genetic divergence followed by population convergence and admixture. Genetic admixture has profound implications for human evolution as it results in the creation of evolutionarily novel genomes that contain combinations of genetic variants (haplotypes) never seen before on the same genomic background. Many of these ancestry-specific variants may have an outsized effect on health-related phenotypes. In addition, when pre-selected genetic variants from distinct parts of the world are brought together in the same admixed population, they can provide raw material for rapid adaptation to the new environment. My PhD thesis research is focused on understanding the implications of genetic admixture for human health, fitness, and population structure. The health and evolutionary dimensions of genetic admixture were interrogated via the analysis of modern, cosmopolitan populations in Latin America.

Latin America is an ideal setting to study the implications of genetic admixture for human health and evolution. Modern human populations in Latin America were formed as part of the Columbian Exchange, starting ~500 years ago when African, European, and Native American populations that were isolated for tens of thousands of years were brought

back together over a relatively short period. The Columbian Exchange represents one of the most abrupt and massive admixture events in human history. The conquest and colonization of the New World entailed the migration of tens of millions of Europeans, the forced migration of more than twelve million enslaved Africans, and tragically, the death of untold millions of Native Americans. Modern Latin American populations emerged from this historical crucible, with the descendants of distinct ancestral groups containing varying degrees of African, European, and Native American ancestry. As such, admixed Latin American genomes contain combinations of ancestry-specific haplotypes never before seen together on the same genomic background.

This dissertation explores the implications of large-scale genetic admixture in Latin America for human health, evolution (natural selection), and population structure (assortative mating). Human health and evolution are explored through the lens of admixture, with an emphasis on the demographic processes that serve to combine distinct ancestry components within genomes. Population structure is considered with respect to assortative mating, which serves to limit the extent of genetic admixture within populations, thereby maintaining genetic diversity among distinct population groups even when they are co-located. In order to understand the implications of admixture for the formation of the New World, comparative genomic analyses were used to characterize patterns of genetic ancestry and admixture for individuals from four modern Latin American populations: Colombia, Mexico, Peru, and Puerto Rico. Genetic ancestry, characterized at both genome-wide (global) and gene (local) levels, was integrated with a variety of functional genomic data sources in an effort to more fully understand the

biological implications of the Columbian Exchange for admixed human populations in the Americas.

Research advance 1: My first study explored the implications of admixture for the genomic determinants of health in admixed Latin American populations. Ancestry patterns for individual genetic variants were characterized for each population, and variants with anomalous ancestry patterns, i.e. present at higher frequency than expected based on the genomic ancestry background, were inspected for their impact on various health-related phenotypes. The four Latin American populations shared evidence of ancestry enrichment for a number of health-related pathways, including cytochrome P450 metabolism, the JAK-STAT signaling pathway, and the Leishmaniasis disease pathway. Variants with an excess of African or European ancestry were also found to be associated with ancestry-specific gene expression. These ancestry-enriched variants were located in genes of the adaptive and innate immune systems and were shown to have population-specific regulatory effects.

Research advance 2: My second study explored the role of admixture in enabling rapid, adaptive evolution in Latin American populations. Genetic ancestry was calculated at the gene level within populations, and gene-specific ancestry deviations, i.e. more or less ancestry than expected given the genomic background, were used to identify signals of admixture-enabled selection. This study relied on a novel combined evidence approach whereby signals of ancestry-enrichment were evaluated for (i) individual genes across multiple populations and (ii) multiple genes that function together to encode polygenic phenotypes. Cross-population signals of African ancestry enrichment were found at the major histocompatibility complex (MHC) locus on chromosome 6, along with concurrent signals of positive selection in ancestral African populations, consistent with admixture-

enabled selection on the adaptive immune system. In addition, a number of phenotypes related to inflammation and both the adaptive and innate immune systems showed evidence for polygenic admixture-enabled selection across the Latin American populations.

Research advance 3: My third study explored the implications of genetic ancestry for the maintenance of human population structure in Latin America. Assortative mating is a universal feature of human behavior, with individuals often choosing mates similar to themselves, and it can serve as a reproductive barrier that maintains structure among distinct population groups even when they are co-located. The exact phenotypic cues and underlying genetic architecture of this phenomenon are unknown; however, it has been shown that genetic ancestry influences assortative mating. In this study, an integrated analysis of local genetic ancestry and the genetic architecture of traits thought to play a role in assortative mating was performed. As assortative mating creates an excess of homozygosity, ancestry-based assortative mating would generate an excess of local ancestry homozygosity. The patterns of local genetic ancestry were utilized to assess the extent of assortative mating for polygenic phenotypes. A novel test statistic, the Assortative Mating Index (AMI) was created to assess the levels of ancestry homozygosity for each polygenic trait. Signals of assortative mating were uncovered in each admixed Latin American population for a number of anthropometric and neurological traits, as well as in the MHC locus.

CHAPTER 1. A BRIEF INTRODUCTION

Here, I provide a brief overview of the major themes that motivate my PhD thesis research. I elaborate on these themes, and provide a more thorough literature review, in Chapter 2.

Anatomically modern humans (AMH) evolved in Africa ~250,000-300,000 years ago [1, 2]. After thousands of years of living on the African continent, AMH began the out-of-Africa migration, ~75,000-100,000 years ago [1, 2]. These early global migrations to Oceania, Europe, Asia, and eventually the Americas resulted in populations that were geographically isolated for tens of thousands of years. During this period, geographic separation of populations was accompanied by reproductive isolation and genetic diversification. Much this divergence can be attributed to genetic drift, whereby allele frequencies fluctuate randomly with no appreciable effect on fitness. However, isolated human populations also adapted to their local environments via positive (diversifying) selection on beneficial alleles. Both genetic drift and positive selection contributed to the genetic diversification of reproductively isolated human populations, with millennia of isolation and divergence giving rise to the major continental population groups recognized today, including Africans, Europeans, and Native Americans.

In the modern era of relatively rapid global travel, the process of migration and subsequent isolation has been reversed. Populations that were isolated for tens of thousands of years are now coming together at an ever-increasing rate. A canonical example of early global travel is the Columbian Exchange, the widespread movement of plants, animals, pathogens, and humans across the Atlantic Ocean to the New World, begun

by Christopher Columbus ~500 years ago [3, 4]. The Columbian Exchange had a profound impact on the trajectory of human evolution: once isolated and genetically diverged continental population group were quickly brought together in a new environment in a way that forever changed the course of human evolution.

This major transition in human global migration had profound implications for many aspects of human biology and evolution, including impacts on health and disease, fitness and adaptation, and population structure. Prior to the Columbian Exchange, distinct African, European, and Native American continental population groups diverged from one another and adapted to their local environments; after the Columbian Exchange, novel admixed populations were formed as combinations of these three ancestry components. This pattern of three-way continental admixture is most pronounced in Latin America. The formation of novel Latin American populations in the New World entailed the combination of numerous pre-adapted haplotypes, allowing for rapid adaptation to the local environment. In modern societies, these admixed populations require different approaches to healthcare and treatment of diseases, as the ancestry-variant relationship and prevalence of diseases differ between populations depending on their admixture patterns [5, 6]. Nevertheless, the majority of studies focusing on understanding the genetic-disease relationship focus on populations with single ancestries, for technical reasons, and findings from these studies are not necessarily portable to those with different or mixed ancestries [7, 8]. Accordingly, while studies of admixed populations pose technical challenges, they also provide unique opportunities for uncovering the genetic architecture of health and fitness.

With respect to the formation of modern, cosmopolitan American populations, two distinct phenomena are of particular interest: 1) admixture – which leads to the breakdown of previously established population structure and 2) assortative mating – which allows for the maintenance of previously established population structure. Admixture occurs when previously isolated populations come together and interbreed, creating individuals with mixed genetic ancestries. Assortative mating, on the other hand, is the process whereby individuals that are more similar choose to mate, yielding more coherent ancestry within groups and more divergent ancestry between groups. These two demographic processes have distinct, yet complementary, implications for human populations and can be considered as two sides of the same coin.

1.1 Admixture

Admixture, the creation of individuals with multiple ancestries due to population mixture, occurs when populations meet in a new environment. Such is the case with the Columbian Exchange in the New World, which resulted in three-way admixture in Latin American populations – individuals with African, European, and Native American genomic ancestry. The population structure found among distinct ancestral source populations is broken down in newly admixed populations. This structure is the result of thousands of years of genetic drift and/or adaptation to local environments, e.g. Malaria resistance in West Africa [9, 10], lactose tolerance in Europe [11, 12] and East Africa [13], and altitude adaptation in Asia [14-16] and South America [16]. As the ancestries were combined in Latin America, ancestry-specific adaptations that were also beneficial in the new environment were introduced at intermediate frequencies, allowing for a rapid population growth should selection act upon them once again [17].

1.2 Assortative mating

Assortative mating, or non-random mating, is a ubiquitous feature of human behavior that affects population genetic structure and evolution [18-20]. It is well known that humans mate assortatively rather than randomly. Mate choices based on similarities in levels of genetic ancestry help to maintain the structure of the ancestral population in the admixed population [18-20]. Human mate choice is often based on traits that another person can see or experience, such as height. As many of these phenotypes vary in their appearance between different populations, assortative mating in admixed populations helps to keep ancestral population structure.

1.3 Admixture versus assortative mating in Latin America

This thesis explores the biological implications of admixture and assortative mating in modern Latin American populations. As mentioned previously, admixture and assortative mating can be viewed as two sides of the same coin, with respect to the implications of the Columbian Exchange that brought together previously isolated continental population groups on a massive scale over a relatively short period of time. Admixture breaks down previously established population structure among ancestral source populations, whereas assortative mating can maintain some degree of structure even among groups that are physically co-located. Both of these distinct demographic processes have implications for health and fitness in modern populations.

Latin American populations represent an ideal system to study the implications of admixture and assortative mating. The modern Latin American populations were formed by genetic admixture among highly divergent ancestral source populations from Africa,

Europe, and the Americas and are distinguished by different levels of continental ancestry from each of these groups and overall admixture. The high levels of genetic divergence among the ancestral source populations that admixed to form modern Latin American populations allow for unambiguous characterization of the ancestral origins of admixed genomes across various levels of scale: genome-wide (global), gene-by-gene (local), and even down to the level of individual genetic variants in some cases. Finally, there are numerous whole genome sequences available from different Latin American populations, which along with thousands of genomes from ancestral source populations provide a rich opportunity for a deep bioinformatic interrogation of these issues.

CHAPTER 2. RAPID, ADAPTIVE HUMAN EVOLUTION FACILITATED BY ADMIXTURE IN THE AMERICAS

2.1 Abstract

Humans have migrated from their ancestral homelands in Africa to nearly every part of the world. Human migration is characterized by a recurrent process of physical isolation and genetic diversification followed by admixture, whereby previously isolated populations come together and exchange genes. Admixture results in the introgression of alleles from ancestral source populations into hybrid admixed populations, and we demonstrate how introgression can facilitate rapid adaptive evolution by introducing beneficial alleles at intermediate frequencies. We provide examples of adaptive introgression between archaic and modern human populations and for admixed populations in the Americas, which were formed relatively recently via admixture among African, European, and Indigenous American ancestral populations. Adaptive introgression has had an outsized effect on the human immune system. In light of the ubiquity of admixture in human evolution, we propose that adaptive introgression is a fundamentally important mechanism for driving rapid adaptive evolution in human populations.

2.2 Human migration, genetic divergence, and admixture

The story of human evolution is one of nearly constant migration. The impulse to leave one's home, explore, and settle new territories is a seemingly universal hominid trait, manifest across multiple species and sub-species of the genus *Homo*, and one that ultimately allowed for humans to populate nearly every corner of the globe. Our hominid ancestors, and their earliest human descendants, have embarked on numerous long distance migrations around the world since their origins on the African continent. Fossil evidence

suggests that *Homo erectus* migrated out of Africa to Eurasia just over 2 million years ago (ya), and *H. heidelbergensis*, a putative ancestor to both archaic and modern humans, left Africa ~800,000 ya [21, 22]. The modern human sub-species – *Homo sapiens sapiens* – is thought to have originated from an African lineage of *H. heidelbergensis* ~300,000 ya [23], and began to migrate out of Africa and around the world starting ~75,000 ya [24, 25]. The phenomenon of migration has had a profound impact on the genetic composition of human populations worldwide, simultaneously driving the joint processes of genetic divergence and admixture. We are particularly interested in how admixture, and the resulting introgression of alleles from ancestral source populations, may have accelerated adaptive evolution in human populations.

H. sapiens' long and steady march out of Africa, through Asia, Oceania, and Europe, and finally throughout the Americas, entailed repeated episodes of population divergence followed by admixture, whereby previously separated populations came together and mixed (Figure 1A). Indeed, we can consider human evolution to be characterized by a recurrent pattern of: (1) (e) migration, (2) isolation, (3) divergence, (4) (im)migration, and (5) admixture (Figure 1B). Human populations constantly migrate to new lands, often resulting in physical isolation, which in turns leads to genetic diversification of the isolated populations. Genetic divergence of isolated populations can occur via genetic drift, owing to small population sizes, and/or natural selection based on local adaptations. However, population isolation does not last forever; eventually, additional waves of migration bring previously isolated populations together again. Any time previously isolated human populations encounter one another, even when those populations are from distinct sub-species, as was the case when modern humans encountered Neandertals and Denisovans,

they interbreed, exchanging genes and yielding new hybrid lineages with genetic contributions from multiple ancestral source populations. As David Reich has detailed in his recent book-length treatment of the ancient DNA revolution, this pattern of migration-driven divergence and admixture has been repeated countless times in archaic and modern human populations around the world [26]. Admixture is not a bug of human evolution – it is in fact a ubiquitous feature of our species [27].

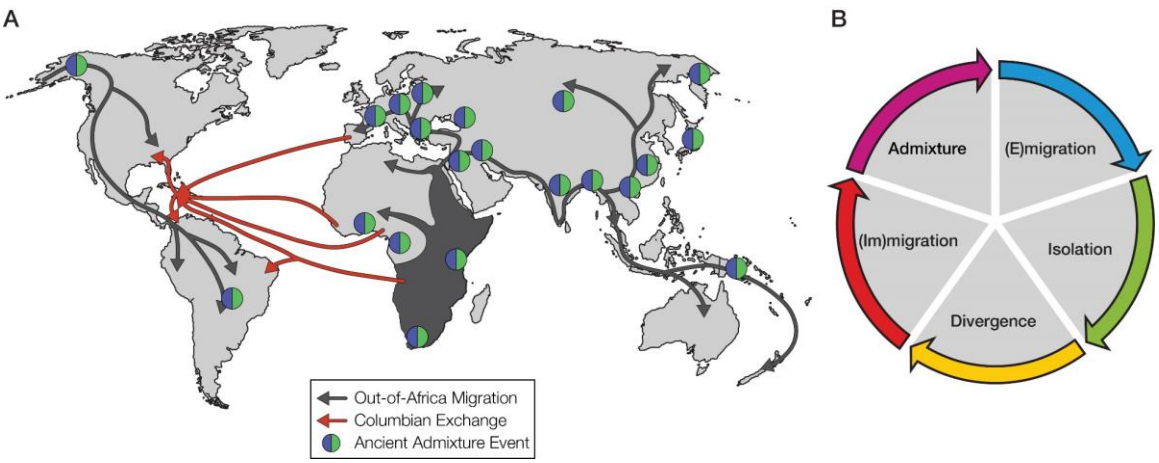

Figure 1. Human migration, genetic divergence, and admixture.

(A) Out-of-Africa migration routes around the world are indicated with gray lines/arrows. Ancient admixture events are indicated with circles [locations taken from [26]], and the modern admixture event that brought together African, European, and Indigenous American populations, i.e. the Columbian Exchange [3, 4], is shown with red lines/arrows. (B) The recurrent and joint processes of migration-driven genetic divergence and admixture are illustrated.

We posit that the process of genetic admixture has had a major impact on accelerating human adaptive evolution, in particular by stimulating the rapid evolution of introgressed alleles in recently admixed populations [17]. This model of human evolution shares much in common with Sewall Wright's Shifting Balance Theory, which emphasized the importance of population sub-division and subsequent migration in facilitating adaptive

evolution [28]. In this chapter, we briefly review studies related to the pace of human adaptive evolution, explaining how admixture can speed up this process, followed by a more detailed treatment of adaptive introgression in both archaic and modern human populations. We emphasize rapid adaptive introgression in the Americas, the region of the world that has experienced perhaps the greatest single admixture event in human history, whereby African, European, and Indigenous American populations that were previously isolated for tens-of-thousands of years were suddenly brought together again following Columbus's arrival in the New World [3, 4].

2.3 Admixture and the pace of adaptive evolution in human populations

In their book *The 10,000 Year Explosion*, authors Cochran and Harpending take aim at the anthropological doctrine which holds that human biological evolution came to a halt around 50,000 ya, thereafter being superseded by a far more dynamic cultural evolution [29]. In distilling this so-called 'conventional wisdom' regarding human evolution, they quote Stephen Jay Gould as saying "There's been no biological change in humans in 40,000 or 50,000 years. Everything we call culture and civilization we've built with the same body and brain". The basic idea underlying this assertion is that the explosion of human culture and behavioral modernity that marked the Upper Paleolithic essentially liberated humans from the strictures of biological evolution. This happened because rapid cultural evolution, in the form of tool and technology development, obviated the need to respond to environmental pressures by the slower process of natural selection. The authors convincingly dismiss with this (perhaps slightly straw man) argument and stress instead that technological developments, the invention of agriculture in particular, actually

accelerated human adaptive evolution by allowing for larger population sizes and consequently more adaptive mutations [30].

The recent acceleration of human adaptive evolution covered in *The 10,000 Year Explosion* was based on the authors' own research along with the work of many other scientists who have taken advantage of the accumulation of human genome sequence variation data to detect signals of adaptive evolution genome-wide in multiple populations around the world [31-34]. This impressive body of research has leveraged the ongoing growth of human population genomic datasets, along with the development of increasingly sensitive methods for detecting adaptive evolution, to steadily decrease the amount of elapsed time needed to observe adaptation events. For instance, the agricultural revolution 10,000 ya led to adaptive evolution for calcium absorption in European populations [35]. Adaptive mutations that conferred lactose tolerance in Europeans [12] and increased energy metabolism in East Asia [36] emerged independently 8,000 ya. Lighter skin pigmentation and increased height were selected for in Europeans 6,000 ya and 5,000 ya, respectively [37, 38]. Sickle cell mutations for protection against malaria were initially proposed to have arisen multiple times in African populations, with estimates around 3,000 ya [9, 10], but a recent study has proposed that these haplotypes are derived from a common ancestral haplotype that emerged 7,300 ya [39]. Perhaps the most recent sequence-based evidence for human adaptive evolution, at the lactase and major histocompatibility loci, dates to 2,000 ya [37]. Here, we present evidence in support of our thesis that adaptive human evolution has occurred in the Americas within the last 500 years, an exceedingly short amount of time with respect to human evolution, via introgression of beneficial haplotypes from ancestral source populations.

Despite the findings on selection outlined in the previous paragraph, human adaptive evolution is still largely regarded as a slow process, which is constrained by the introduction of new adaptive alleles via mutation. This can be illustrated by the classic population genetic model showing the rate at which the frequency of an adaptive allele will increase in a population (Figure 2). A new mutant allele will be introduced at the low population frequency of $1/(2 \times N_e)$, where N_e is the effective population size. For example, a relatively small effective population size of 5,000 will yield an initial mutant allele frequency of 0.01%. If the new allele is adaptive, selection will act to increase its population frequency (p) over time proportional to the selection (s) and dominance (h) coefficients, according to the recursion equation $p_{i,t+1} = p_{i,t}w_i/\overline{w}$, where i is the allele, t is the current generation, p is the allele frequency, w_i is the marginal fitness of the allele i, and $\overline{w}$ is the population mean fitness. The increase in adaptive allele frequency happens extremely slowly at the end of the low end of the allele frequency spectrum. Under an additive dominance model ($h = 0.75$) with a strong selection coefficient of $s = 0.1$, it will take more than 100 generations to see a 20% increase in the initial frequency of the adaptive allele. However, as can be seen in Figure 2, the rate of adaptive allele frequency increase speeds up tremendously at intermediate allele frequencies. Under the same dominance and selection parameters, but starting from an intermediate allele frequency of 35%, a doubling of the adaptive allele frequency can occur in less than 20 generations. This feature of adaptive evolution is what leads us to believe that admixture can facilitate extremely rapid human adaptation. When two or more previously isolated populations converge and mix, they introduce alleles to the newly formed admixed population at intermediate frequencies proportional to the percent contributions of each ancestral population. Since admixture

introduces new alleles at intermediate frequencies via introgression in this way, it has the potential to allow for substantial increases in the frequency of adaptive alleles over a relatively small number of generations.

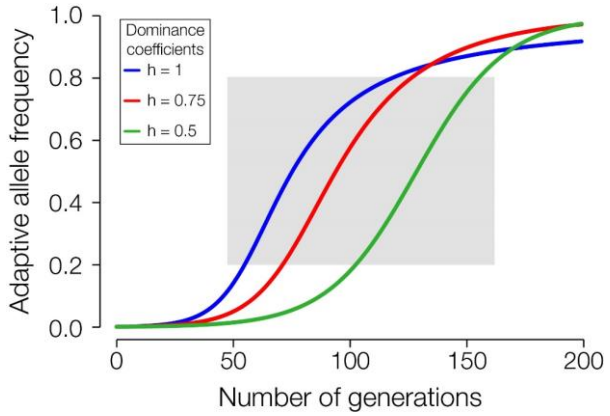

Figure 2. Population genetic model showing the increase in frequency of an adaptive allele over time.

Adaptive allele frequencies are shown on the y-axis, and time in generations is shown on the x-axis. The model corresponds to a selection coefficient (s) of 0.1, and three different allele increase trajectories are shown according to different dominance coefficients (h). The gray shading corresponds to the zone at intermediate allele frequencies where the rate of adaptive allele frequency increase is most rapid.

While our notion that admixture can dramatically speed up adaptive human evolution is based on a theoretical conjecture, there are numerous empirically observed cases of adaptive introgression in human populations that support this view of evolution. A majority of the studies on adaptive introgression focus on the impacts of admixture between modern humans and archaic hominids, *e.g.* Neandertals and Denisovans. In the sections that follow, we first discuss a few examples of ancient adaptive introgression, as they are by now more widely accepted. Then we review an example of more recent adaptive introgression in Africans. Finally, we discuss adaptive introgression in the Americas, where there are fewer studies and more contention regarding the results. Findings from

admixed American populations are particularly provocative in the sense that they point to adaptive evolution occurring over a span of 500 years, or approximately 20-25 generations, an exceptionally short period of time for human evolution.

2.4 Ancient adaptive introgression

Evidence of adaptive introgression acting as a means of rapid human adaptive evolution has been shown via the admixture of modern humans with archaic hominids. Before the modern human out-of-Africa migration, Europe and Asia were populated by other *Homo* species: Neandertals and Denisovans. These archaic populations had isolated and genetically diverged from the ancestors of modern humans hundreds of thousands of years before and were adapted to their respective local environments in Europe and Asia. As modern humans emerged from Africa, they not only encountered new environments, but also these new human-like populations. Admixture of archaic and modern humans created new genomes that combined intermediate frequency adaptive alleles from the archaic populations into the genomic background of the modern humans. This is thought to have helped modern human populations to more quickly adapt to new environments as they settled Europe and Asia, including fighting off novel pathogens. Here we will review a few examples of ancient adaptive introgression; for a more comprehensive review, see [40].

2.4.1 Immune system

A key factor in the response of the immune system is the *HLA* class I genes of the major histocompatibility complex (MHC). These genes are involved in antigen presentation for immune cell recognition and are very diverse across human populations.

Abi-Rached et al. were interested in the evolutionary source of the deeply divergent human *HLA-B* allele, *HLA-B*73:01* [41]. This allele is found mainly in west Asia and is in linkage with *HLA-C*15:05*, found in west and southeast Asia. The authors suggested that this allele combination was not present in Africa prior to the out-of-Africa migration. To test this hypothesis, they characterized the archaic *HLA* class I genes from Denisovans and Neandertals, finding that two *HLA-A* and two *HLA-C* alleles in Denisovans were most similar to modern sequences, including the *HLA-C*15* allele. Geographic locations of these similar modern alleles show a very low presence in Africa, with higher presence in Asia and Oceania. Divergence estimates of these alleles show that they were formed before the out-of-Africa migration, but since the alleles are not present in Africa, they are likely to come from an archaic source populations, namely Denisovans. The authors completed the same analysis with similar Neandertal *HLA-A* and *HLA-C* alleles and found them to be present in Eurasians and absent in Africans. These findings suggest that the immune systems of archaic *Homo* populations were better adapted to local pathogens in the Eurasian environments, and upon migrating out of Africa, modern humans rapidly adapted the immune system through introgression of these sequences.

After the study on *HLA* adaptive introgression, Mendez et al. interrogated *STAT2*, an immune system gene with a role in interferon-mediated responses, for signals of adaptive introgression with Neandertals in Europeans [42]. Initial sequencing of *STAT2* revealed the 'N' haplotype, a deeply divergent haplotype present only in the non-African populations used in this study. When compared with the draft Neandertal reference sequence, the N haplotype most closely matched the Neandertal sequence. In addition to this, the haplotype linkage disequilibrium (LD) including *STAT2* is very long in both West

Eurasians (~130 kb) and East Asians and Melanesians (~260 kb), a finding expected if the haplotype was introduced via introgression. The upper divergence estimate between the Neandertal and N haplotype of ~160 kya is more recent than the estimates of divergence between the Neandertal and human reference sequence of ~600 kya. The N haplotype has a 10-fold higher frequency in Melanesians, particularly the long form of the haplotype, when compared with the rest of the Eurasian populations. This suggests that *STAT2*, or some other gene in the haplotype, was the target of positive selection in Melanesians. In either case, the introgression of *STAT2* into the modern human genome affected interferon signaling in the immune system.

Genome-wide studies on Neandertal and Denisovan introgression had identified a number of putative Neandertal introgressed regions in modern humans [43, 44]. Dannemann et al. utilized these genomic maps to characterize the adaptive introgression potential of a region containing a haplotype of three Toll-like receptors (TLRs), *TLR6-TLR1-TLR10*, which show some of the highest probabilities of Neandertal introgression [45]. These genes play an important role in the innate immune system as they are the first line of defense against pathogens and help to activate the adaptive immune response. To characterize this region of chromosome 4 as resulting from an adaptive introgression event, the authors identified evidence of introgression by showing that there are seven main haplotypes for the TLRs and three of these are more similar to archaic sequences than to the rest of modern humans as determined by sequence comparisons. In addition to this, the geographic distribution of the archaic-like sequences provide evidence for introgression as they are mostly found outside of Africa, which is to be expected if introgression occurred when modern humans moved into Neandertal and Denisovan environments. A high

differentiation, as determined by F_{ST}, and previous studies reporting signatures of positive selection on SNPs in *TLR10* provided the authors with evidence that the haplotype was under positive selection. Expression quantitative trait loci analysis showed tissue-specific regulatory effects increasing the expression of these genes in white blood cells. Overall, it was posited that this TLR haplotype could have been adaptively introgressed into the modern human population as a means to affect the innate immune system response with response to potential pathogens, including *H. pylori*. The authors mention that a diversity of haplotypes for TLRs could have increased the adaptation of modern humans in novel environments, such as that encountered after migration out of Africa.

2.4.2 Integumentary system

Adaptive introgression of Neandertal sequences have also affected the hair and skin phenotypes of Europeans and East Asians, as published by two studies in 2014. Vernot and Akey, as mentioned previously, created a catalog of putative introgressed Neandertal sequences in 379 Europeans and 298 East Asians, part of the Phase 1 data release of the 1000 Genomes Project, using a modified S^* summary statistic to identify signals of introgression followed by sequence comparison to a Neandertal reference [44]. The authors then interrogated these sequences to find significantly differentiated introgressed regions between Europe and East Asia as well as shared regions in the two populations with relatively high allele frequency. Using F_{ST} to identify differentiated variants between the two populations, the authors identified two regions with genes that are part of the integumentary system, in addition to other regions and genes. *BCN2* on chromosome 9 was found to be at ~70% frequency in Europeans while absent in East Asians; this gene has been related to skin pigmentation in Europeans. *POU2F3* on chromosome 11 was

found to be at ~66% frequency in East Asians and <1% in Europeans; this gene is expressed in the epidermis and mediates keratinocyte proliferation and differentiation. Both populations shared a total of six regions with >40% allele frequency with signals of adaptive introgression. One of these regions on chromosome 12q13 contains a type II cluster of keratin genes, providing evidence for adaptive variation of skin phenotypes.

A second study to find evidence of adaptive introgression in the integumentary system used a conditional random field approach to identify putative Neandertal introgressed regions in 1,004 modern humans from the Phase I data release of the 1000 Genomes Project [43]. They also created a map of Neandertal introgression events in the modern human genome and analyzed the top 5% of genes with the highest inferred Neandertal ancestry. They found that there was a significant enrichment of genes involved in keratin filament formation and posit that Neandertal variants may have helped modern humans adapt to the novel European and Asian environments by affecting their skin and hair.

2.4.3 Altitude adaptation

A well-studied example of how archaic introgression can provide an adaptive benefit to a modern human population is that of altitude adaptation in Tibetans. Early studies elucidated the molecular underpinnings of adaptation to high altitude living in Tibetans and the high-altitude Sherpa population and noted that *EPAS1* and *EGLN1* are crucial to controlling signals of hemoglobin concentration in Tibetans, compared with other lowland East Asian populations [46-48]. Jeong et al. [15] determined that the Tibetan population was a result of an ancestral “high-altitude” population admixing with a “low-altitude”

population. As *EPAS1* had the strongest signal of selection for Tibetans, Huerta-Sanchez et al. [14] moved forward to identify the source of the adaptive introgression for this gene. After re-sequencing 40 Tibetan individuals (high-altitude) and 40 Han Chinese individuals (low-altitude), the authors found that F_{ST} for this gene was highly differentiated as expected if there was selection on the high-altitude haplotype in Tibetans. When comparing the Tibetan-specific haplotype to potential donor sequences, it was determined that the haplotype shared more sequence similarity with the Denisovan haplotype than any other extant or archaic population. The authors concluded that introgression from a Denisovan population allowed the modern Tibetan population to rapidly adapt to the high-altitude of the Tibetan plateau.

2.5 Adaptive introgression in modern humans

The Bantu-speaking populations (BSPs) of Africa experienced multiple periods of adaptive introgression during their migration throughout Africa over the last 1,500 years [49]. As the BSPs migrated, they encountered other African populations, already adapted to their local environments. Genetic ancestry characterization of the BSPs showed admixture with other African populations, including western rainforest hunter-gatherers, eastern African farmers, and San populations. In each of the BSP populations – western, eastern and southeastern – the authors searched for evidence of excess ancestry and compared their findings with signals of positive selection to identify putative adaptively introgressed regions. In western BSPs, there was an overlap of excess western rainforest hunter-gatherer ancestry with a strong signal of positive selection for the *HLA* region and a moderately strong signal for *CD36*. These regions are both related to the immune system response, with *CD36* being associated with susceptibility to malaria caused by *Plasmodium*

falciparum. In eastern BSPs, excess eastern African ancestry overlapped a moderately strong signal of positive selection for the lactase gene (*LCT*) providing the lactase persistence phenotype for eastern BSPs. Thus, both the immune system and diet-related phenotypes of BSPs have been influenced by recent admixture and adaptive introgression in these African populations.

2.6 Adaptive introgression in the Americas

We are particularly interested in studying adaptive introgression in admixed American populations [17]. Modern (cosmopolitan) human populations in the Americas were formed primarily by admixture among ancestral source populations from Africa, Europe, and the Americas (Figure 1A). This is considered to be one of the largest and most abrupt admixture events in all of human evolution and one that has had a profound effect on world history [3, 4]. Admixed American genomes can be considered as evolutionarily novel in the sense that they contain combinations of ancestry-specific alleles that never previously existed together on the same genomic background. The creation of such novel admixed genome sequences has important implications for health and fitness in modern American populations [50]. The possibility of adaptive introgression in the Americas can be considered controversial in light of the fact that the ~20-25 generations that have elapsed since the process of admixture in the Americas began represents a very short amount of time in terms of human evolution, less than 1% of the time that has elapsed since humans migrated out of Africa. In principle, it should be very difficult to observe adaptive human evolution over such a short time scale. Nevertheless, we contend that the Americas represent an ideal laboratory to study adaptive introgression and to explore the possibility of extremely rapid adaptation in human populations.

Our working hypothesis is that numerous alleles were 'pre-selected' in ancestral source populations over thousands of years based on their utility in the local environments of Africa, Europe, and the Americas. Subsequently, when the ancestral source populations were suddenly brought back together, some of these pre-selected alleles could have also provided an adaptive benefit in the New World environment. For example, alleles that served to protect their human hosts against infectious disease may have been particularly important in the pathogen-rich environment of the Americas. In addition, neutral alleles that diverged in frequency among ancestral populations based on genetic drift could later become adaptively beneficial in the new environment. In either case, adaptively beneficial alleles introduced at intermediate frequencies via introgression could quickly increase in frequency owing to their utility in the novel admixed populations (see Figure 2 and the previous discussion on how introgression can accelerate adaptive evolution).

The analytical approach used to test this hypothesis is based on the delineation of genome-wide 'local ancestry' patterns in admixed populations. Local ancestry refers to the specific ancestral origins – African, European, or Native American – for specific chromosomal regions (*i.e.* haplotypes). Local ancestry is assigned by comparing individual chromosomal segments against the corresponding genomic regions from panels of ancestral population reference genomes [51, 52]. Once this is done for an admixed American population, the population's local ancestry fractions for any given region of the genome can be compared to the genome-wide population ancestry averages to search for 'ancestry-enriched' regions (Figure 3). Ancestry-enriched regions are genomic segments that show anomalously low or high ancestry fractions, for any given ancestry component, compared to the genome-wide population averages. Statistically significant local ancestry

deviations are taken to represent evidence of adaptive introgression. The presence of independent signals of previous positive selection on these same regions in ancestral source populations can be used to provide additional evidence in support of adaptive introgression [49]. Below, we review all currently known cases of adaptive evolution in admixed American populations.

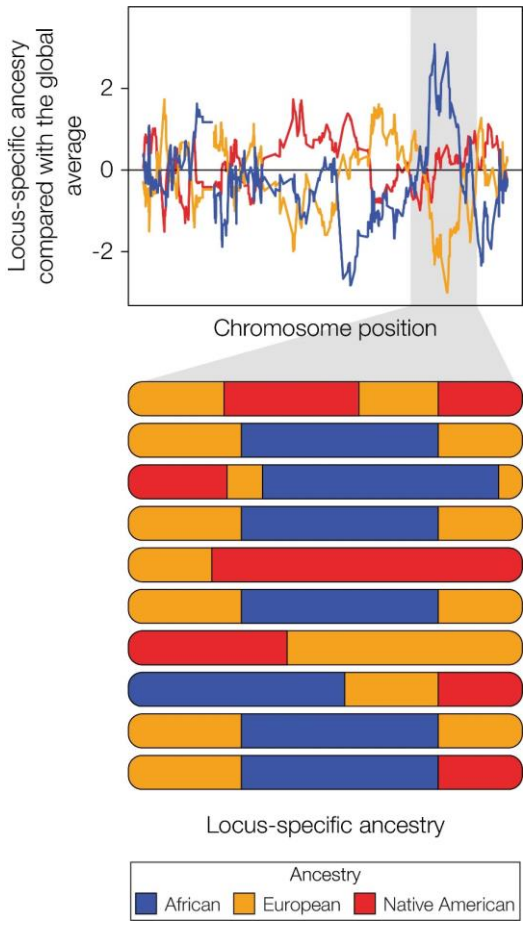

Figure 3. Ancestry-enrichment analysis for identifying adaptive introgression events.

Locus-specific ancestry patterns – African (blue), European (orange), and Native American (red) – are characterized across all chromosomes in the population. Locus-specific ancestry fractions are compared with the global ancestry fractions along the genome to identify ancestry-enriched (or depleted) segments. An example is shown (gray shading) for a region that is enriched for African ancestry and depleted for European ancestry.

2.6.1 Puerto Rico

One of the first studies on adaptive introgression in the Americas uncovered signals of recent selection and introgression in Puerto Ricans [53]. Genetic ancestry characterization of the study population – 192 Puerto Ricans as part of the Genetics of Asthma in Latino Americans study – showed that the population was mainly of European descent with relatively equal fractions of African and Native American ancestries. Local ancestry estimates, when compared with global averages, revealed excess African ancestry on chromosome 6 and excess Native American ancestry on chromosomes 8 and 11. The region of chromosome 6 with African enrichment harbors the major histocompatibility complex (MHC), the first response to invading pathogens for the adaptive immune system. Chromosome 11 shows Native American enrichment of an olfactory gene cluster, which the authors mention, has been shown to be under positive selection and reference other studies [54, 55]. The final Native American enrichment on chromosome 8 did not have any clear candidates for adaptive introgression, but the one gene in the region, *CSMD3*, did have some tissue-specific expression. The authors suggest that the enrichment of African alleles of the immune system could be due to them being more advantageous in the new environment as they are genetically more diverse than Europeans or Native Americans and better able to fight off the numerous pathogens imported and endemic to the area. Another reason for the ancestry enrichment could be that these genes played a role in a phenotype that offered a fitness advantage and thus was acted upon by natural selection upon admixture in the New World. For either reason, the immune system of Puerto Rican individuals seems to have been influenced by adaptive introgression of ancestry-specific alleles.

2.6.2 Colombia

A 2015 study to characterize the genetic ancestry of the Colombian population from the 1000 Genomes Project found that similar to Puerto Rico, the Colombian population has mostly European ancestry, followed by Native American and finally African ancestry fractions [56]. In addition to understanding the impacts of sex-biased admixture, the authors evaluated local ancestry patterns, looking for locus-specific patterns using a trinomial probability metric to signal areas with excess ancestry. The regions with anomalous ancestry patterns were then interrogated for previously identified signs of positive selection and their phenotypic associations, specifically health-related phenotypes. The authors found signals of African ancestry enrichment for *HLA-B* on chromosome 6, part of the MHC. There was also evidence of enrichment of European ancestry for *SLC24A5* on chromosome 15, a gene that influences skin pigmentation with decreased melanin. In addition to single genes with anomalous ancestry patterns, the authors performed a gene set enrichment analysis to identify pathways containing an abundance of ancestry-enriched genes. They found both the innate and adaptive immune response contained signs of ancestry enrichment. These lines of evidence provide support for the hypothesis that the Colombian genomes retained adaptive ancestry-specific loci that were better suited to combat the wide variety of pathogens in the environment.

As a caveat to the findings, it should be noted that the program used for local ancestry assignment generated very short ancestry tracts, compared with the tract lengths generated from current local ancestry assignment programs such as RFMix [51]. It is unlikely that in the ~20-25 generations since the Columbian Exchange that there would be

sufficient genomic recombination to generate the size tracts seen in this analysis and their findings should be validated using contemporary programs.

2.6.3 Mexico

In 2014, a novel two-layer hidden Markov model was proposed for inferring local ancestry of admixed individuals based on the detection of haplotype structure [57]. This method was used to characterize the genetic ancestry and highlight regions with excess ancestry-specific loci in Mexican individuals from the HapMap3 Project and 1000 Genomes Project. For both projects, the author found excess African ancestry on chromosome 6 in the MHC region, as well as on chromosome 8p23.1 at a known chromosomal inversion.

A follow-up study on recent selection in Mexican individuals found a significant excess of African ancestry of the MHC on chromosome 6 [58]. The authors analyzed genetic ancestry data from two cohorts of Mexican individuals and found excess ancestry in both cohorts. By being able to replicate the results in a second cohort, there was more confidence in the findings of African ancestry enrichment. In addition to this, the authors developed a technique to infer local ancestry without the concern of inaccurate Native American samples and estimated the amount of selection necessary for this locus to remain significantly enriched with African ancestry. The results suggest that selection was at similar strengths to that of lactase selection in Europeans and the sickle-cell trait in Africans. The authors suggest that the challenging conditions of the New World, with the existing pathogens and those brought over by the Spaniards and Africans could have provided the necessary selection pressure for the African MHC alleles to increase in

frequency due to the greater diversity of African MHC. Infectious diseases and epidemics, in addition to harsh living conditions, caused many of the Native American populations to perish, leading to a lack of Native American ancestry at this important immune system locus.

2.6.4 African Americans

A study on ancestry-specific selection in 1,890 African Americans characterized the genetic ancestry of the population to be ~72% African and ~28% European [59]. The authors then identified loci with European (or African) ancestry 3 standard deviations above or below the genome-wide average as those potentially under natural selection. Four regions were found to have excess European ancestry and two were found to have excess African ancestry. The excess European ancestry was found to be related to response to influenza infection and the African ancestry was found to be related to general immune system signaling. Using F_{ST} between African Americans and putative ancestral African populations, the authors identified signals of positive selection and found four regions that carried differentiated SNPs, including those related to malaria response and the MHC, suggesting the immune system of African Americans in the New World was under pressure with response to the novel pathogen environment. It is important to note, however, that the regions with excess African ancestry do not overlap with those putatively under positive selection.

In response to these findings, another group performed a genetic ancestry analysis on 29,141 African American individuals and found there to be no evidence of directional selection [60]. Instead, the authors state that the earlier findings could have been due to

chance or systematic biases in data handling. When directly comparing the 2012 results to the authors' calculations, the authors showed that they get much lower ancestry deviation estimates than previously reported, and the 2012 enrichment findings were not replicated. The same results are found when focusing on the positive selection estimates using F_{ST}; the previous results were not replicated and there were no overlaps between the two findings.

2.7 Conclusions and future prospects

Adaptive human evolution was postulated to have stopped ~40-50,000 ya, coincident with the emergence of cultural evolution [61, 62]. Not only has this anthropological doctrine been shown to be false, but there is abundant evidence that adaptive evolution has actually accelerated over the last 10,000 years [29]. Throughout this chapter, we have provided examples of adaptive introgression acting as a means for facilitating rapid human evolution, through admixture with Neandertals and Denisovans thousands of years ago, for admixture in modern Africans 1,500 years ago, and via admixture in the Americas 500 years ago. Admixture and introgression have allowed modern humans to colonize new environments, such as the Tibetan plateau, resist novel pathogens and combinations of pathogens throughout Europe, Africa and the Americas, and generally be better suited to their local environments. The immune system has been shown to be a hotspot of adaptive introgression throughout time as admixture among previously adapted populations allowed modern humans to quickly adapt and thrive in their new environments.

As seen in the contradictory studies of African Americans, signals of adaptive introgression can be due to biases in the data or small sample sizes. To increase confidence

in ancestry enrichment acting as signals of adaptive introgression, future studies can ensure that regions with ancestry enrichment also have multiple lines of evidence of positive selection from ancestral populations on the same genomic regions in the admixed populations. In addition, if there are multiple admixed populations with similar ancestral source populations, finding signals of ancestry enrichment and positive selection shared among the multiple populations could provide more confidence that a finding is true as there is likely shared selection pressures in the same area of the world. Selection is also likely to occur on phenotypes which are caused by multiple genes, not just one genomic locus. If the same ancestry enrichment and positive selection signals are seen in multiple genes encoding a polygenic phenotype, then it is likely that adaptive introgression could have led to rapid adaptive human evolution.

Finally, we would like to propose the idea that admixture, now recognized as a ubiquitous feature of human evolution [26, 27], has had a far greater impact on human adaptive evolution than formerly recognized. The linked processes of genetic divergence followed by admixture (Figure 1B) have provided abundant raw material for adaptive introgression throughout human evolutionary history, which in turn can provide for extremely rapid adaptive evolution unconstrained by mutation. It may well be the case that, owing to admixture, adaptive evolution of human populations is far more common and much more dynamic than previously imagined.

CHAPTER 3. GENETIC ANCESTRY, ADMIXTURE AND HEALTH DETERMINANTS IN LATIN AMERICA

3.1 Abstract

3.1.1 Background

Modern Latin American populations were formed via genetic admixture among ancestral source populations from Africa, the Americas and Europe. We are interested in studying how combinations of genetic ancestry in admixed Latin American populations may impact genomic determinants of health and disease. For this study, we characterized the impact of ancestry and admixture on genetic variants that underlie health- and disease-related phenotypes in population genomic samples from Colombia, Mexico, Peru, and Puerto Rico.

3.1.2 Results

We analyzed a total of 347 admixed Latin American genomes along with 1,102 putative ancestral source genomes from Africans, Europeans, and Native Americans. We characterized the genetic ancestry, relatedness, and admixture patterns for each of the admixed Latin American genomes, finding a spectrum of ancestry proportions within and between populations. We then identified single nucleotide polymorphisms (SNPs) with anomalous ancestry-enrichment patterns, *i.e.* SNPs that exist in any given Latin American population at a higher frequency than expected based on the population's genetic ancestry profile. For this set of ancestry-enriched SNPs, we inspected their phenotypic impact on disease, metabolism, and the immune system. All four of the Latin American populations

show ancestry-enrichment for a number of shared pathways, yielding evidence of similar selection pressures on these populations during their evolution. For example, all four populations show ancestry-enriched SNPs in multiple genes from immune system pathways, such as the cytokine receptor interaction, T cell receptor signaling, and antigen presentation pathways. We also found SNPs with excess African or European ancestry that are associated with ancestry-specific gene expression patterns and play crucial roles in the immune system and infectious disease responses. Genes from both the innate and adaptive immune system were found to be regulated by ancestry-enriched SNPs with population-specific regulatory effects.

3.1.3 Conclusions

Ancestry-enriched SNPs in Latin American populations have a substantial effect on health- and disease-related phenotypes. The concordant impact observed for same phenotypes across populations points to a process of adaptive introgression, whereby ancestry-enriched SNPs with specific functional utility appear to have been retained in modern populations by virtue of their effects on health and fitness.

3.2 Background

The modern human species – *Homo sapiens sapiens* – originated in sub-Saharan Africa ~200,000 years ago and began to migrate out of Africa and spread throughout the world starting ~70,000 years ago [1, 2]. After heading north out of Africa, humans spread to the east and west, populating Melanesia, Asia, and Europe, and eventually made their way across the Bering Strait into the Americas ~20,000 years ago. As human populations occupied different parts of the globe, they often became geographically isolated in their

new homelands. Thousands of years of geographic isolation were accompanied by population genetic diversification, giving rise to the diverse human population groups that can be seen around the world to this day [63, 64]. Distinct continental population groups – African, Asian, and European in particular – are the most obvious examples of this evolutionary process. There were, of course, a number of episodes of genetic admixture during this time [27], whereby previously isolated populations came into contact and began to mix, but for the most part, the dominant mode of human evolution since our emergence from Africa has been characterized by populations' physical isolation followed by genetic diversification.

This pattern of human evolution was turned upside down upon the arrival of Columbus in the New World a mere 500 years ago, which is less than 1% of the elapsed time since humans emerged from Africa. Columbus' voyages precipitated the so-called 'Columbian Exchange' – a massive transfer of life forms, which had evolved separately for millennia, between the Old and New World hemispheres [3, 4]. The human dimension of the Columbian Exchange entailed genetic admixture between previously isolated populations on an unprecedented scale, in terms of both scope and rapidity [17]. The conquest and colonization of the Americas, along with the trans-Atlantic slave trade, brought African, European, and Native American populations into close and sustained contact for the first time. As a consequence, these diverse population groups began to mix, giving rise to novel admixed American populations. This is particularly true for Latin America, where populations are characterized by high levels of genetic admixture among African, European, and Native American ancestral source populations [65-67].

Latin American genomes can thus be considered to represent a recent innovation in human evolution. Indeed, genomes from modern Latin American populations are evolutionarily novel in the sense that they contain combinations of genetic variants (haplotypes) that never previously existed together on the same genetic background. Our group is interested in trying to understand the implications of the recent advent of novel Latin American genomes, particularly as it relates to the genetic determinants of health-related phenotypes. In other words, we are asking what it means when genomes that were separated for many thousands of years are suddenly brought back together and what the implications of this process are for human health and fitness.

Our group and others have employed an approach that we call ancestry-enrichment analysis to address these kinds of questions via population-level studies of admixed American genomes [17]. This approach relies on the characterization of local patterns of genetic ancestry for individual genomic loci. Local ancestry assignment, colloquially referred to as chromosome painting, entails the delineation of ancestral origins of specific haplotypes across the genome. The resulting chromosome paintings reveal the genomes of admixed individuals as mosaics of interspersed ancestry-specific haplotypes. When a population sample of admixed genomes is characterized in this way, the percent ancestry contributions from each ancestral source population can be computed for all haplotype loci genome-wide. Ancestry-enrichment analysis then entails the identification of specific haplotype loci that have anomalous patterns of local ancestry, *i.e.* levels of locus-specific ancestry that are significantly higher or lower than can be expected by chance given the overall ancestry profile of the population. Statistically significant signals of ancestry-enrichment are taken as evidence of adaptive introgression, whereby introgressed

haplotypes increase in frequency by virtue of a selective advantage that they provide to individuals in an admixed population.

A number of recent studies have used ancestry-enrichment analysis to show evidence of adaptive introgression in admixed American genomes. The first study of this kind showed an excess of African ancestry at the major histocompatibility locus (MHC) in a sample of Puerto Rican genomes [53], and a follow up study several years later also found ancestry-enrichment at the same region in a Mexican population [57]. Since that time, several other studies have replicated the finding of ancestry-enrichment in admixed populations at this and other health related loci [56, 58, 59, 68, 69]. Our own more recent work on Colombian genome sequences revealed even more widespread ancestry-enrichment, which impacted various aspects of the immune system, including pathways involved in both innate and adaptive immunity [56].

All of the previous ancestry-enrichment studies were distinguished by their interrogation of the ancestral origins of individual haplotypes, *i.e.* physically linked sets of genetic variants that are inherited together. For this study, we developed and applied a novel method based on individual genetic variants – single nucleotide polymorphisms (SNPs) – in an effort to expand our view of the relationship between genetic ancestry, admixture and health in Latin American populations. Our approach entails the detection of SNPs that are found at anomalously high frequencies in admixed populations compared to what is expected based on their frequencies in the ancestral source populations, *i.e.* ancestry-enriched SNPs. To find such ancestry-enriched SNPs, we consider the proportional contributions of ancestral source populations to admixed Latin American populations, together with SNP frequencies in the ancestral populations, to derive expected

SNP frequencies for the Latin American populations. These expected frequencies are compared to observed frequencies in order to identify statistically significant ancestry-enriched SNPs; the connection between ancestry-enriched SNPs and health-related phenotypes is then explored via analysis of the functional annotations of the SNPs and their linked genes. In particular, we interrogated the impact of ancestry-enriched SNPs on disease, metabolism and immune system pathways. This approach uncovered signals of ancestry-enrichment in health- and disease-related traits shared among all four of the Latin American populations that we analyzed, raising the possibility of shared selective pressures among them.

3.3 Materials and Methods

3.3.1 Comparative genomic data sources

Whole genome sequences from four admixed Latin American populations – Colombia (*n*=94), Mexico (*n*=64), Peru (*n*=85), and Puerto Rico (*n*=104) – were taken from the 1000 Genomes Project (1KGP) phase 3 data release[1] [70]. Genome sequences from the four Latin American populations were compared to whole genome sequences and whole genome genotypes of global reference populations from African, European, and Native American continental population groups to characterize their patterns of genetic ancestry and admixture (Figure 4 and Table 1). The global reference whole genome sequences were also taken from the 1KGP, and reference whole genome genotype data was taken from the Human Genome Diversity Project (HGDP) [64]. The whole genome sequence and whole genome genotype data were merged, with sites that existed in both

[1] http://www.internationalgenome.org/data/

datasets retained for subsequent analysis, and PLINK v1.9 [71] was used to correct single nucleotide polymorphism (SNP) strand orientation as needed. This resulted in a dataset of 435,782 SNPs from 1,449 individuals, across 19 populations. The final merged SNP dataset was phased with the 1KGP haplotype reference panel using the program SHAPEIT2 [72].

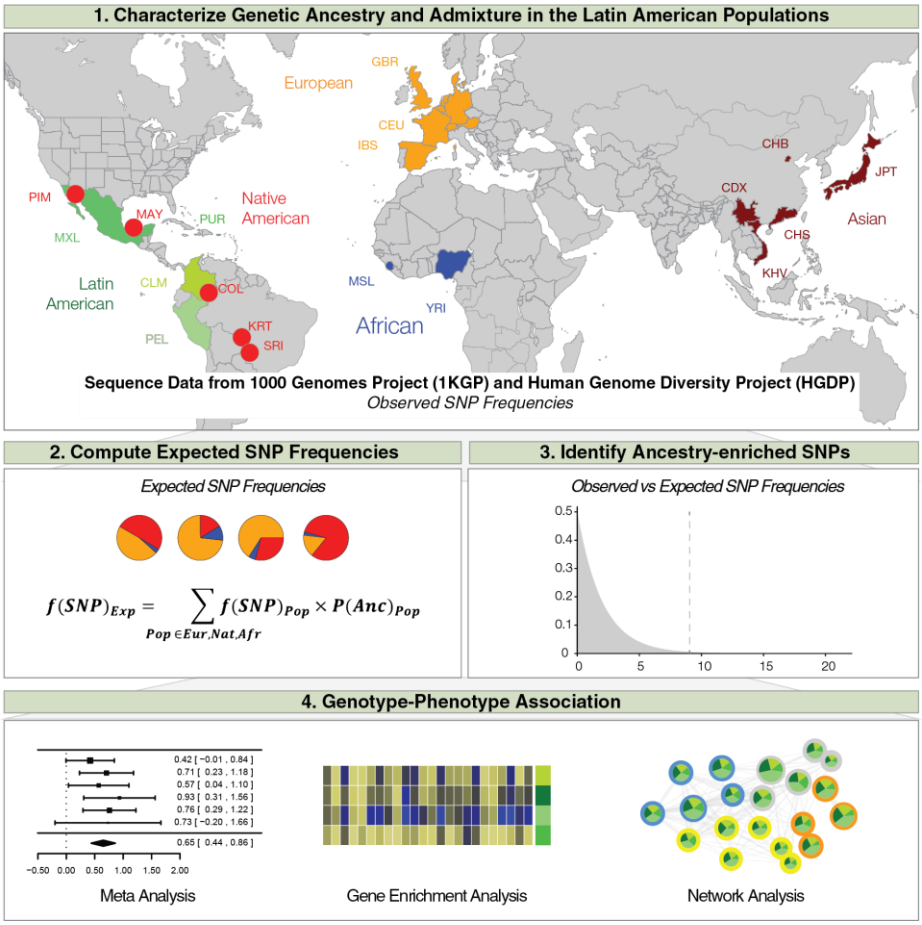

Figure 4. Analysis scheme used for this study.

(1) Genetic ancestry and admixture profiles were characterized for four Latin American populations. (2) Expected SNP frequencies in the admixed Latin American populations are calculated based on their ancestry profiles. (3) Ancestry-enriched SNPs are identified by comparing observed versus expected SNP allele frequencies in in the admixed Latin American populations. (4) Ancestry-enriched SNPs are mapped to genes, which in turn are used for gene set enrichment in order to identify impacted health-related pathways and phenotypes.

Table 1. Human populations analyzed in this study.

Populations are organized into continental groups: African, East Asian, European, and Admixed American from the 1000 Genomes Project (1KGP) and Native American from the Human Genome Diversity Project (HGDP). Short names, descriptions, and the numbers of genomes analyzed (n) are provided for each individual population.

Dataset	Short	Full Description	*n*
1 KGP African (n=193)	MSL	Mende in Sierra Leone	85
	YRI	Yoruba in Ibadan, Nigeria	108
1KGP East Asian (n=504)	CDX	Chinese Dai in Xishuangbanna, China	93
	CHB	Han Chinese in Bejing, China	103
	CHS	Southern Han Chinese, China	105
	JPT	Japanese in Tokyo, Japan	104
	KHV	Kinh in Ho Chi Minh City, Vietnam	99
1KGP European (n=297)	CEU	Utah residents with NW European ancestry	99
	IBS	Iberian populations in Spain	107
	GBR	British in England and Scotland	91
1KGP Admixed American (n=347)	CLM	Colombian in Medellin, Colombia	94
	MXL	Mexican Ancestry in Los Angeles, California	64
	PEL	Peruvian in Lima, Peru	85
	PUR	Puerto Rican in Puerto Rico	104
HGDP Native American (n=108)	KRT	Karitiana in Brazil	24
	SRI	Surui in Brazil	21
	COL	Colombians in Colombia	13
	MAY	Maya in Mexico	25
	PIM	Pima in Mexico	25

3.3.2 Genome ancestry assignment

ADMIXTURE [73] was run on individuals from both the global reference and admixed Latin American populations to infer their genome-wide ancestry profiles. ADMIXTURE was run with K=3 corresponding the three ancestral continental population groups: African, European, and Native American. The ADMIXTURE results for the admixed Latin American populations were used to infer individuals' percent ancestry contributions from each of these three continental ancestry groups. The program RFMix [51] was used to assign the continental ancestry origins of individual haplotypes across the

genome, *i.e.* local ancestry. As with ADMIXTURE, African, European, and Native American populations were used as reference populations for RFMix. Ancestry-specific haplotypes were only called for regions where RFMix certainty was at least 99%.

3.3.3 Detection of ancestry-enriched SNPs

Ancestry-enriched SNPs were characterized as SNPs found in higher frequencies in admixed Latin American populations compared to what is expected based on (1) their frequencies in the ancestral source populations, and (2) the proportion of ancestry derived from each ancestral source population. For any given SNP, in any given Latin American population, the expected frequency of the SNP $f(SNP)_{Exp}$ can be calculated as:

$$f(SNP)_{Exp} = \sum_{Pop \in Eur,Afr,Nat} f(SNP)_{Pop} \times P(Anc)_{Pop} \quad (1)$$

where $f(SNP)_{Pop}$ is the frequency of the SNP in a specific ancestral source population and $P(Anc)_{Pop}$ is the proportion of ancestry in the modern Latin American population derived from that same ancestral population. Ancestry proportions were computed using the reduced, merged set of SNPs described above, with African and European reference populations from the 1KGP and Native American reference populations from HGDP. SNP frequencies were computed using whole genome sequences, in order to provide more complete coverage of variants genome-wide, using the YRI African and IBS European reference populations with the most closely related East Asian population CHB taken as a surrogate for Native American ancestry.

The statistical significance of SNP ancestry-enrichment calculated this way was determined by comparing the observed (Obs) to expected (Exp) frequencies of the reference (Ref) and alternate (Alt) alleles for any given SNP as shown here:

$$\chi^2 = \frac{\left(Obs_{Ref} - Exp_{Ref}\right)^2}{Exp_{Ref}} + \frac{(Obs_{Alt} - Exp_{Alt})^2}{Exp_{Alt}} \quad (2)$$

The χ^2 distribution was used to calculate *P*-values for each SNP, and false discovery rate (FDR) *q*-values were used to account for multiple statistical tests. The SNPs that had significant FDR values ($q < 0.05$) were considered to be ancestry-enriched in the Latin American population, *i.e.* present at a higher frequency than expected based on the population ancestry profile.

For ancestry-enriched SNPs, the individual ancestry components (Anc) that gave rise to the pattern of enrichment were also determined by jointly minimizing the frequency difference between the SNP in the Latin American population and a single ancestral source population while maximizing the distance between that single source population and the other two ancestral populations:

$$Anc_{Pop1} = \left[\left(f(SNP)_{Pop1} - f(SNP)_{Pop2}\right) + \left(f(SNP)_{Pop1} - f(SNP)_{Pop3}\right)\right]/2 - \left|f(SNP)_{Obs} - f(SNP)_{Pop1}\right| \quad (3)$$

where $f(SNP)_{Popx}$ is the frequency of the SNP in each of the ancestral source populations and $f(SNP)_{Obs}$ is the frequency of the ancestry-enriched SNP in the Latin American population.

3.3.4 Gene set enrichment analysis

Ancestry-enriched SNPs were mapped to genes if they mapped within the NCBI RefSeq [74] gene models, *i.e.* between transcription start and stop sites, on the UCSC Genome Browser human genome reference sequence build GRCh37/hg19. Functionally coherent gene sets were curated from the Molecular Signatures Database (MSigDB) version 5.1 [75] for three broad functional categories: health- and disease-related phenotypes, metabolism, and immunity. Gene set enrichment analysis (GSEA) was performed for each Latin American population by adopting the MSigDB statistical framework to find functional gene sets that were enriched for genes with mapped ancestry-enriched SNPs. To do this, genes that harbor ancestry-enriched SNPs were overlapped with genes from each functional gene set, and overlap enrichment was performed using the R limma package [76]. Overlap enrichment *P*-values were computed for each gene set using the Wilcoxon signed-rank test.

3.3.5 Expression quantitative trait loci (eQTL) analysis

RNA-seq data were taken from the GUEVADIS RNA sequencing (RNA-seq) project for 1KGP samples[2]. A total of 445 RNA-seq samples were used in the analysis, including 87 African and 358 European individuals. The RNA-seq data correspond to gene expression levels for the same lymphoblastoid cell lines, *i.e.* Epstein–Barr virus (EBV) transformed B-lymphocytes, which were used for the 1KGP DNA-seq characterization. RNA-seq sample preparation, sequencing experiments and read-to-genome mapping were

[2]ftp://ftp.ebi.ac.uk/pub/databases/microarray/data/experiment/GEUV/E-GEUV-1/analysis_results/

performed as previously described [77]. The read-to-genome mapping corresponds to human genome build GRCh37/hg19. Gene expression levels were quantified based on RNA-seq mapped reads and corresponded to ENSEMBL gene models [78]. Gene expression levels were quantified by using the reads per kilobase per million mapped reads (RPKM) approach in combination with the probabilistic estimation of expression residuals (PEER) method as previously described [79, 80].

Matched whole genome sequencing based SNP genotype calls were obtained for the same 445 individuals from the phase 3 release of the 1KGP [70]. Only SNPs with minor allele frequency (MAF) greater than 5% were used for the downstream analysis to ensure both the confidence of genotype calls and the reliability of the eQTL association analyses. Gene expression levels of 445 individuals were then regressed against their SNP genotypes to identify eQTLs using the program Matrix eQTL [81]. The Matrix eQTL program was run using the additive linear model option with gender and population labels included as covariates.

3.3.6 SNP pathway meta-analysis

Meta-analysis was used to evaluate the overall ancestry-enrichment for sets of SNPs that are implicated in specific health- or disease-related pathways. For any given ancestry-enriched SNP that mapped to a gene found in an overrepresented pathway, a log odds ratio was calculated as:

$$Log\ Odds\ Ratio = ln\left[\frac{f(Obs)_{Ref}/f(Exp)_{Ref}}{f(Obs)_{Alt}/f(Exp)_{Alt}}\right] \quad (4)$$

where the observed (Obs) versus expected (Exp) frequencies (f) are compared for the SNP reference (Ref) versus alternate (Alt) alleles. Then for each pathway, the set of individual SNP log odds ratios was combined to yield pathway-specific log odds ratio values along with 95% confidence intervals using the fixed-effect model with moderators via linear (mixed-effect) models implemented in the metafor package in R [82].

3.4 Results

3.4.1 Relating genome ancestry and health in Latin America

We developed and applied a single nucleotide polymorphism (SNP)-based approach to relate genome ancestry to genetic determinants of health in admixed Latin American populations (Figure 4). First, patterns of genetic ancestry and admixture in Latin American populations were characterized via comparison with reference genome sequences of putative ancestral source populations from Africa, the Americas and Europe (Table 1). We then computed the expected SNP frequencies in Latin American populations by taking into consideration the SNP frequencies in the ancestral source populations along with the proportional contributions of each ancestral source population to the modern Latin American populations. Comparisons of observed versus expected SNP frequencies in admixed Latin American populations were used to identify what we refer to as 'ancestry-enriched' SNPs, which are SNPs found at anomalous frequencies in Latin American populations compared to what can be expected based on their ancestry profiles. Ancestry-enriched SNPs were mapped to genes, and then genes were used in gene set enrichment analysis to identify impacted health-related pathways and phenotypes.

3.4.2 Genetic ancestry and admixture in four Latin American populations

Genome sequences from four Latin American populations – Colombia, Mexico, Peru, and Puerto Rico – were compared to whole genome sequences and whole genome genotypes of global reference populations from African, European, and Native American continental population groups in order to characterize their patterns of genetic ancestry and admixture. Each Latin American population has a distinct pattern of three-way continental genetic admixture characterized by population-specific proportions of African, European and Native American ancestry (Figure 5). Puerto Rico and Colombia are characterized by relatively high levels of three-way admixture, with substantial ancestry contributions from all three continental population groups, whereas Mexico and Peru have primarily Native American and European ancestry. Puerto Rico and Colombia also have the highest levels of European ancestry, while Peru and Mexico have majority Native American ancestry. The 80% Native American ancestry component for Peru is the single highest contribution of any ancestral population to an admixed Latin American population, and the 2% African ancestry fraction for this same population is the lowest. African source populations contribute the least amount of ancestry to all four Latin American populations analyzed here. The continental ancestry proportions for each Latin American population were used as described in the following section to detect ancestry-enriched SNPs that exist in any given population at a higher frequency than expected based on its ancestry profile.

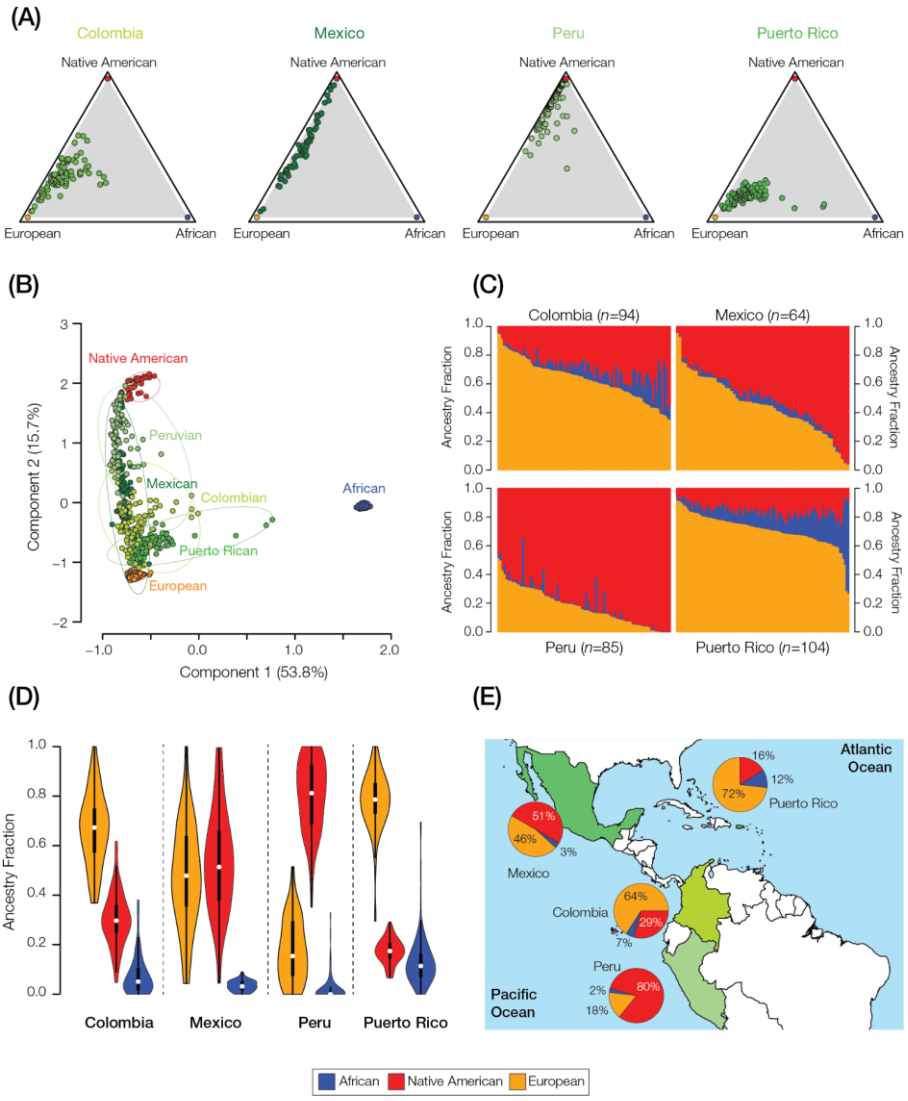

Figure 5. Genetic ancestry and admixture in Latin American populations.

The ancestry contributions of putative ancestral source populations to four modern, admixed Latin American populations are shown. (A) Triangle plots showing the relative ancestry contributions – African, European, Native American – to admixed individuals from four Latin American populations. (B) PCA plot showing the genetic relationships among individuals from admixed Latin American populations compared to putative ancestral source populations. Each population is bounded by a minimum spanning ellipse. (C) Admixture plots showing the fractions of African, Native American and European ancestry among admixed individuals from four Latin American populations. Each individual is represented as a column with the admixture fractions color coded as shown in the legend. (D) Violin plots showing distributions of ancestry fractions among individuals from four Latin American populations. (E) Pie charts showing the average ancestry values for each population next to its geographic location.

3.4.3 Ancestry-enriched SNPs in Latin American populations

Our approach to relating genetic ancestry to determinants of health and disease in modern Latin American populations relies on the detection of SNPs that are found at anomalously high frequencies in admixed populations compared to what is expected based on their frequencies in the ancestral source populations, *i.e.* ancestry-enriched SNPs. We reasoned that such ancestry-enriched SNPs are likely to have an outsized effect on health and disease in modern Latin American populations, perhaps related to an initial increase in population frequency via adaptive introgression.

We developed and applied a quantitative method to identify individual SNPs that are enriched in admixed Latin American populations with respect to ancestry from one of the three ancestral source populations: Africa, Europe, and the Americas. To do so, the expected frequencies for each SNP were calculated using the frequency of the given SNP in each of the three ancestral source populations conditioned upon the proportion of each ancestral source population in the Latin American population of interest. Observed SNP frequencies were compared to expected SNP frequencies to identify ancestry-enriched SNPs; the details of this approach are shown in section *3.3 Materials and Methods*.

The distributions and median values of ancestry-specific SNP χ^2 values are shown in Figure 6A. Peru shows the strongest overall signal of SNP ancestry-enrichment, followed by Mexico, Colombia, and Puerto Rico. Statistically significant ancestry-enriched SNPs for each population were mapped to genes for subsequent analysis of their impact on health- and disease-related phenotypes. There is a substantial amount of overlap of mapped genes among the four populations (Figure 6B). Out of 156 total genes with

mapped ancestry-enriched SNPs, 102 (65%) are shared among two or more populations compared to 54 (35%) that are population-specific. There are 40 genes that bear ancestry-specific SNPs in all four Latin American populations, which is by far the single largest component of shared versus unique genes. Lists of all SNPs that show significant ancestry-enrichment – along with details regarding their observed and expected allele frequencies, test-statistic values, and specific ancestry enrichment patterns – can be found in Table 5.

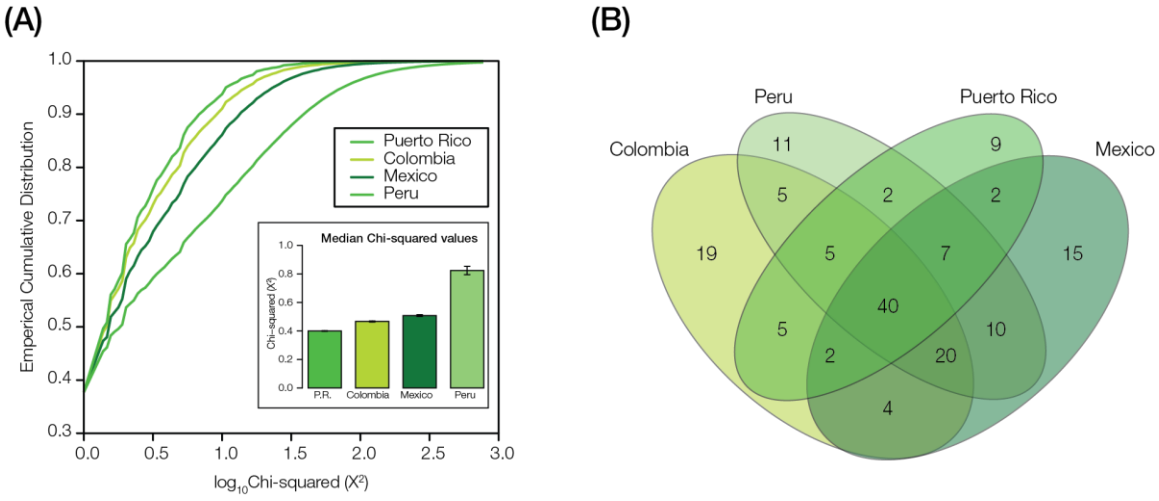

Figure 6. Ancestry-enriched SNPs in Latin American populations.

An overview of the distributions of ancestry-enriched SNPs within and between the four admixed Latin American populations are shown, giving an indication of the overall numbers of ancestry-enriched SNPs along with the extent to which they are shared or unique to specific populations. (A) Cumulative distributions of ancestry-enrichment χ^2 values for all SNPs in the four Latin American populations. Inset: Median χ^2 values for each population ± standard error. (B) Venn diagram showing the number of genes with significant ancestry-enriched SNPs exclusive to one population and those shared by more than one population.

3.4.4 Gene set enrichment analysis of overrepresented SNPs

The genes with mapped ancestry-specific SNPs were analyzed with gene set enrichment analysis (GSEA) to look for overrepresented health- or disease-related pathways and phenotypes (Figure 7). This approach allowed us to identify the specific pathways and phenotypes that are most affected by ancestry-enriched SNPs. The presence of significantly overrepresented pathways and/or phenotypes in two or more populations was taken to indicate a higher likelihood of genetic ancestry effects on health and disease in modern Latin American populations.

A number of pathways and phenotypes have significantly overrepresented ancestry-enriched SNP genes in all four populations (Figure 7A). These include gene sets related to immunity (yellow) and metabolism (orange) as well as several disease-specific gene sets (blue) (Figure 7B). Immune system pathways with ancestry-enriched SNPs include the cytokine receptor interaction, T cell receptor signaling, and antigen processing and presentation pathways. Implicated metabolic pathways include both drug and xenobiotic metabolism as well as steroid hormone biosynthesis. Diseases of note include several pathologies that are known to be found in high prevalence in Latin American populations: type I diabetes, Alzheimer's disease and Leishmaniasis. A number of other signaling pathways were implicated by this analysis – calcium, MAPK, and GNRH signaling – pointing a role for ancestry-enriched SNPs in mediating human-environment interactions. Lists of all pathways that show significant enrichment of genes with mapped ancestry-enriched SNPs, for each admixed Latin American population – along with information regarding the overlapping genes and pathway enrichment statistical significance (FDR *q*-values) – are provided in Table 6.

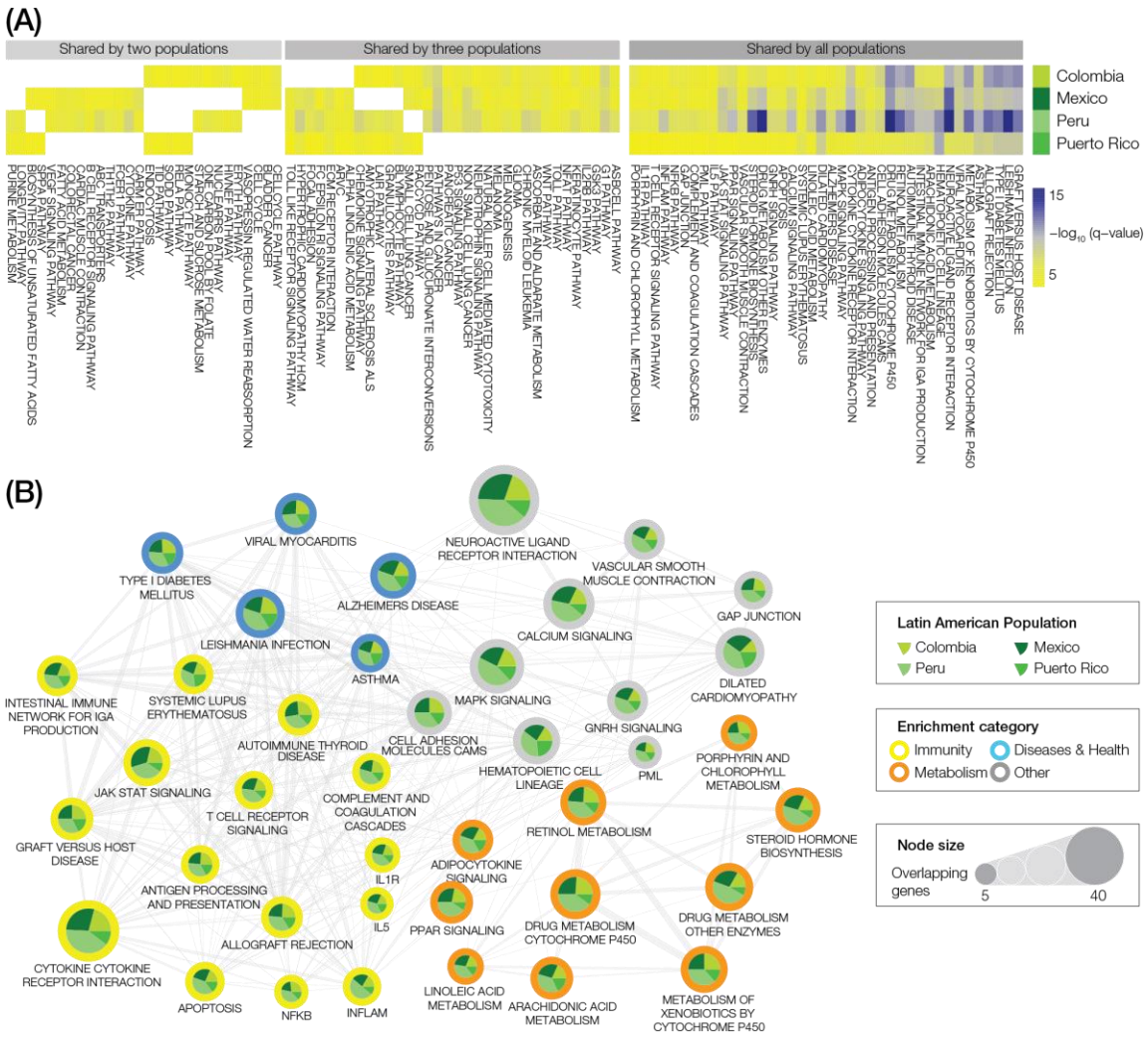

Figure 7. Gene set enrichment analysis of ancestry-enriched SNP genes.

*Functionally coherent gene sets and pathways that are overrepresented with respect to ancestry-enriched SNPs are shown, giving an indication of the kinds of health-related phenotypes that have been shaped by genetic ancestry in the four admixed Latin American populations. (A) Heatmap showing significantly enriched functional gene sets (*i.e., *pathways and phenotypes) shared by two, three or all four Latin American populations. The cells are color coded by the log transformed statistical significance (FDR* q*-value) of gene set enrichment analysis results. (B) Network showing significantly enriched pathways and phenotypes shared by all four Latin American populations. Nodes sizes represent the number of ancestry-enriched genes in each set. Pie charts show how many genes in a given set are from each population. Color coding describes the functional enrichment category as shown in the legend.*

We focused on several notable examples of health- and disease-related pathways that were found to have significantly overrepresented ancestry-enriched SNP genes in all four Latin American populations (Figure 8). For each of these pathways, and in each population studied, we performed additional meta-analyses of the sets of mapped ancestry-enriched SNPs in order to evaluate the pathway's overall ancestry enrichment. We also computed analogous overall observed versus expected allele frequencies for each pathway in all four populations. There are 15 genes from the Leishmaniasis immune response pathway with mapped ancestry-enriched SNPs, including a pair of Toll-like Receptor encoding genes as well as several interleukin genes (Figure 8A). The meta-analysis for this pathway shows an overall ancestry-enrichment for all SNPs in each of the four populations analyzed here. Leishmaniasis is a parasitic disease with high prevalence in the tropics and subtropics including Latin America. Similar pathway-specific analysis revealed overall ancestry-enrichment for SNPs linked to drug metabolism (Figure 8B), including multiple genes from the cytochrome P450 family, as well as the JAK-STAT signaling pathway, which is activated by cytokines as part of the innate immune response (Figure 8C). The ancestry-enrichment observed for the drug metabolism pathway could represent an adaptation based on detoxification linked to local dimensions of diet and environmental exposure in the New World.

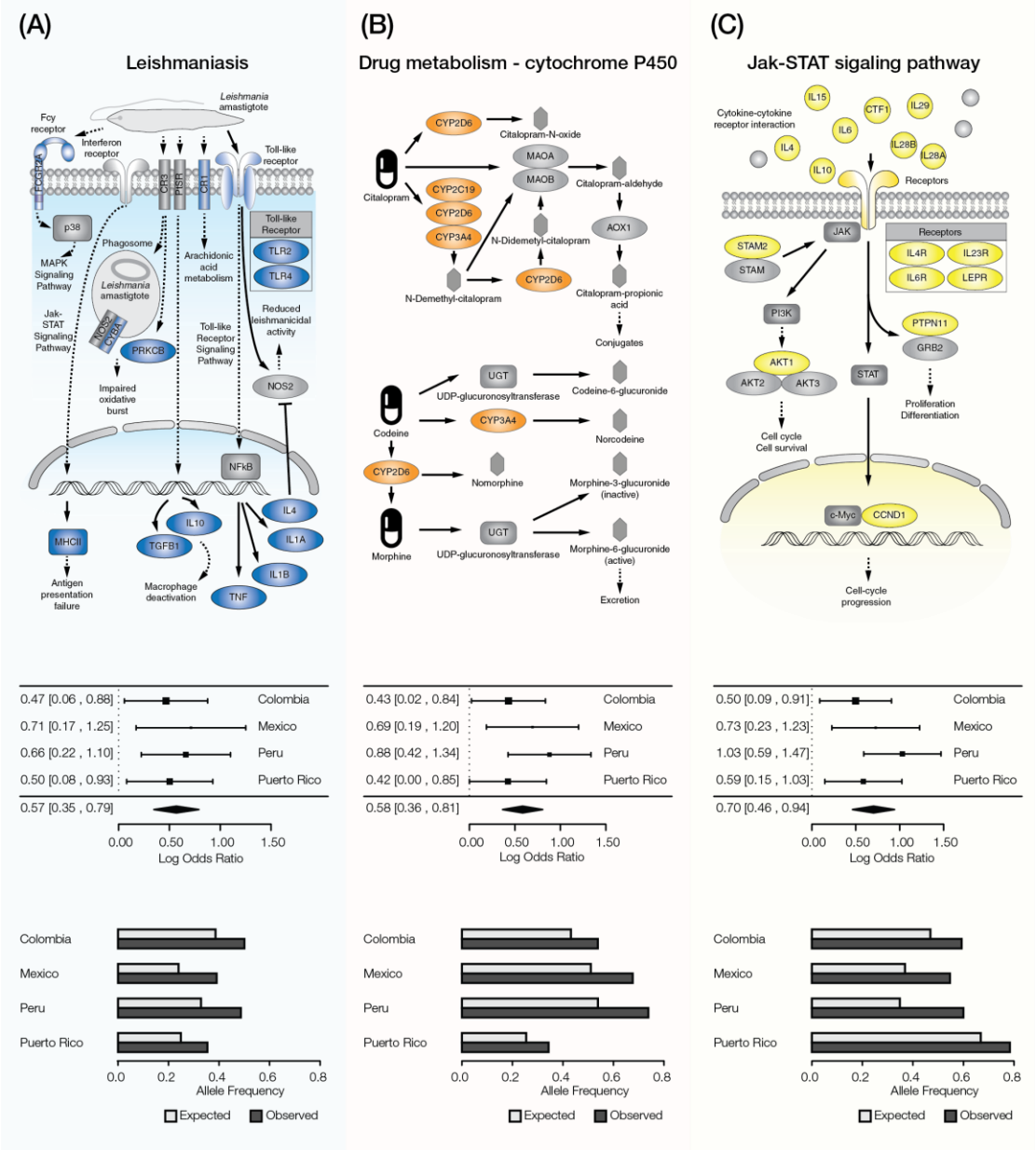

Figure 8. Pathways with ancestry-enriched SNP genes in functional categories of interest.

These results highlight examples of specific health-related functions and pathways that have been shaped by genetic ancestry in the four admixed Latin American populations. For each functional category, a pathway schematic is shown, indicating the pathway genes and their roles, along with meta-analysis results and observed versus expected SNP frequencies for each population. (A) Leishmaniasis, an example of a disease and health related pathway. (B) Cytochrome P450 drug metabolism, an example of a metabolism related pathway. (C) Jak-STAT signaling pathway, an example of an immune-related pathway.

3.4.5 Ancestry-specific expression quantitative trait loci (eQTL)

We explored the effects of ancestry-specific SNPs on gene regulation via expression quantitative trait loci (eQTL) analysis. eQTL are individual SNPs with genotype variants that are associated with gene expression levels; associations of this kind point to a role for SNP variants in gene regulation (*e.g.*, via differential transcription factor binding affinities and/or allele specific expression levels) [83, 84]. To do this, we searched for ancestry-enriched SNPs that have ancestry-specific or shared genotype-expression associations. The first step of this analysis entailed the identification of the specific ancestry-components that predominantly contribute to the observed patterns of SNP ancestry-enrichment (see section *3.3 Materials and Methods*). SNPs with highly asymmetric ancestry-enrichment patterns, *i.e.* predominant contributions from a single ancestral source population, were then chosen for eQTL analysis.

Using this approach, we found a number of cases of SNPs that show overrepresented African or European ancestry in modern Latin American populations and are also associated with ancestry-specific gene regulation (Figure 9). A number of the genes regulated by ancestry-specific SNPs were found to play specific roles in the immune system and infectious disease responses. In particular, genes from both the innate and adaptive immune system were found to be regulated by ancestry-enriched SNPs that exert population-specific regulatory effects (Figure 9A-D). For example, African ancestry-enriched SNPs were found to exert African-specific regulatory control over genes for both immunoglobulin receptors (PVR and TYROBP) and a downstream tyrosine kinase (ZAP70) involved in the adaptive immune response (Figure 9E). Similarly, European ancestry-enriched SNPs were also found to act as population-specific eQTLs with

regulatory effects on that were specific to the European populations. Analogous patterns of ancestry-specific SNP enrichment and gene regulatory control were found for genes involved in cytokine-receptor interactions, hematopoietic cell development, and cell-cell immunomodulatory interactions.

3.5 Discussion

Latin America has a unique genetic heritage with high levels of admixture from African, European, and Native American ancestral source populations [65-67]. As such, the genome sequences of Latinos contain combinations of ancestry-specific genetic variants that never previously existed in the same genomic background. In other words, Latin American genomes represent a very recent evolutionary innovation in the long trajectory of human evolution and migration around the globe. Accordingly, the development and application of genomic approaches to healthcare in Latin America will require a deep understanding of the genetic ancestry and admixture profiles of Latin American populations. This issue is particularly pressing given the fact that the vast majority of studies aimed at uncovering genetic variants associated with health- and disease-related phenotypes have been conducted in populations with European ancestry [7, 8].

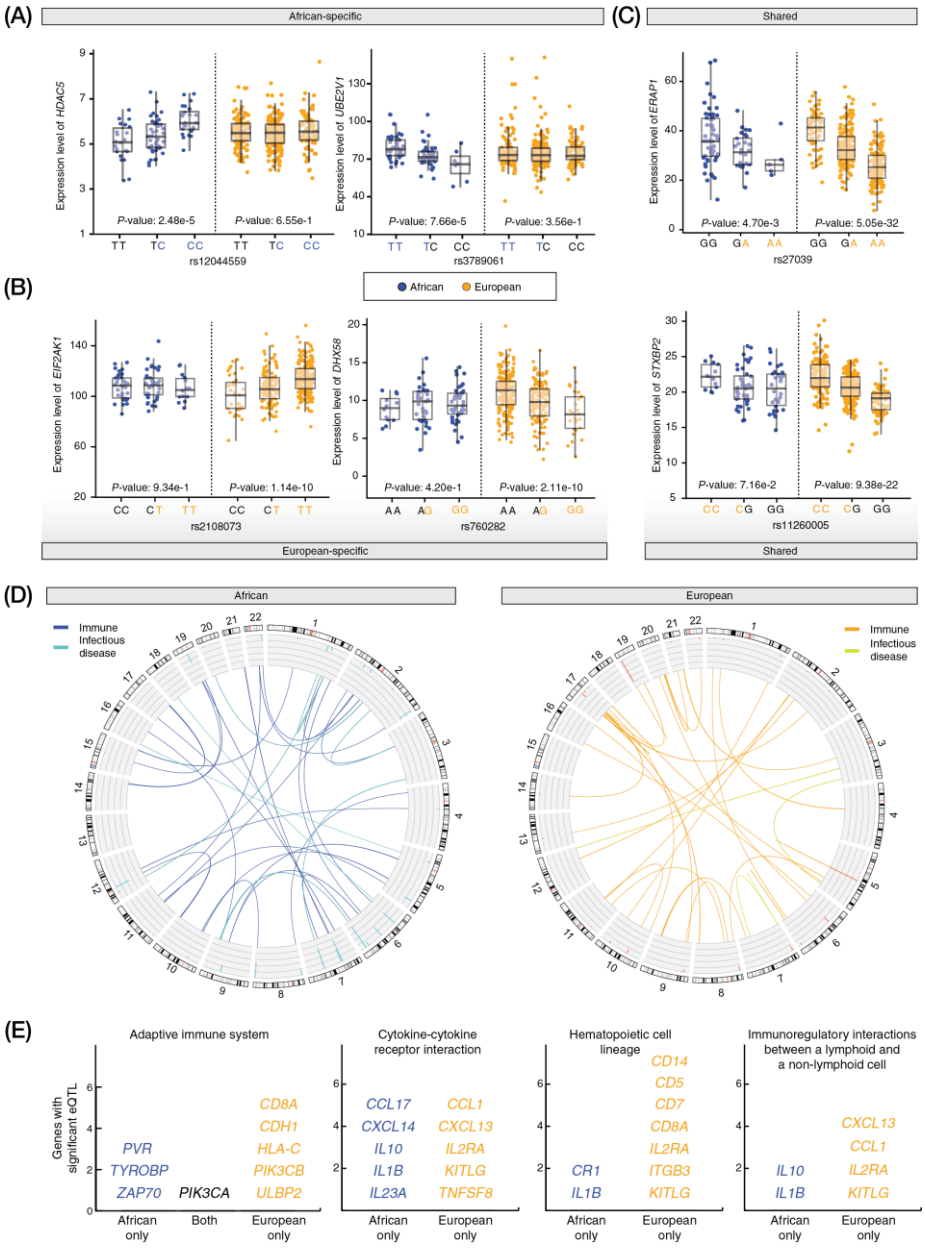

Figure 9. Ancestry-specific effects on gene expression.

These results give an indication of how ancestry-enriched SNPs can impact health-related phenotypes by virtue of their gene regulatory effects. SNP-by-ancestry interactions were characterized using an expression quantitative trait loci (eQTL) approach. Examples of (A) African-specific and (B) European-specific eQTL are shown along with (C) eQTL shared between populations. (D) eQTL related to immune system and infectious disease found in the African and European populations are shown in a CIRCOS plot with links indicated between eQTLs and their regulated genes. (E) Examples of immune-related pathways that include multiple eQTL-regulated genes for African and/or European populations.

Here, we have tried to address this issue by relating patterns of genetic ancestry and admixture to health and disease determinants in Latin American genomes. To do so, we developed and applied a novel SNP-based approach to ancestry enrichment analysis. Our approach leverages information on the genetic ancestry of the modern Latin American populations to discover SNP variants that exist in a given population at higher frequencies than expected, *i.e.* ancestry-enriched SNPs. We found that specific sets of ancestry-enriched genetic variants, from each of the three ancestral source populations, have been preferentially retained in modern Latin American populations based on a variety of roles that they play in health and fitness. These findings have relevance for the development of genomic approaches to healthcare, *i.e.* personalized or precision medicine, in Latin America.

Gene set enrichment analysis uncovered a number of immunity, metabolism, and disease-related pathways that are significantly overrepresented with respect to genes that contain ancestry-enriched SNPs (Figures 7 and 8). These results suggest that these particular pathways, and their related phenotypes, could underlie population-specific health disparities in the four admixed Latin American populations studied here. They also give an indication that populations with particular ancestry profiles may be more or less disposed to some of these diseases and phenotypes; information of this kind could ultimately help to guide targeted health interventions. As these results represent basic research into the relationship between genetic ancestry and determinants of health, more clinically facing (translational) research will need to be done in order to precisely define the role of individual ancestry-enriched variants in disease etiology, prevention and treatment.

Expression quantitative trait loci (eQTL) analysis revealed ancestry-enriched SNPs in modern Latin American populations that are associated with African- or European-specific patterns of gene regulation (Figure 9). This includes SNPs that are associated with ancestry-specific regulation of genes involved in both the innate and adaptive immune systems as well as targeted infectious disease responses. These results underscore the relevance of gene regulatory control as an underlying driver of adaptive introgression in admixed populations.

One important caveat with respect to the interpretation of the results that we report is that they can only be taken to apply to the four specific populations analyzed here: Colombian in Medellin, Colombia (CLM), Mexican Ancestry in Los Angeles, California (MXL), Peruvian in Lima, Peru (PEL), and Puerto Rican in Puerto Rico (PUR). Given the diversity of Latin American populations, and in particular their distinct ancestry profiles, we should expect to see distinct ancestry enrichments for different countries in the region, such as Argentina, Chile, Brazil, etc. This caveat not only applies to different countries but also applies to different populations within the same country. Colombia, for instance, is an extremely diverse country with populations from different regions that show very distinct ancestry profiles [65]. The population of Colombia analyzed here is from Medellín in the state of Antioquia, and this particular population shows averages of 64% European ancestry, 29% Native American ancestry, and 7% African ancestry. However, we have previously shown that the population from the neighboring state of Chocó has a totally distinct ancestry profile with 76% African ancestry, 13% European ancestry, and 11% Native American ancestry [5, 85, 86]. Accordingly, results from the analysis of the population from Medellín cannot be taken to represent the entire country of Colombia.

Clearly, a deeper understanding of the relationship between genetic ancestry and health determinants in Latin America will require analysis of many more populations within and between the region's countries.

CHAPTER 4. ADMIXTURE-ENABLED SELECTION FOR RAPID ADAPTIVE EVOLUTION IN THE AMERICAS

4.1 Abstract

4.1.1 Background

Admixture occurs when previously isolated populations come together and exchange genetic material. We hypothesized that admixture can enable rapid adaptive evolution in human populations by introducing novel genetic variants (haplotypes) at intermediate frequencies, and we tested this hypothesis via the analysis of whole genome sequences sampled from admixed Latin American populations in Colombia, Mexico, Peru, and Puerto Rico.

4.1.2 Results

Our screen for admixture-enabled selection relies on the identification of loci that contain more or less ancestry from a given source population than would be expected given the genome-wide ancestry frequencies. We employed a combined evidence approach to evaluate levels of ancestry enrichment at (1) single loci across multiple populations and (2) multiple loci that function together to encode polygenic traits. We found cross-population signals of African ancestry enrichment at the major histocompatibility locus on chromosome 6, consistent with admixture-enabled selection for enhanced adaptive immune response. Several of the human leukocyte antigen genes at this locus (*HLA-A*, *HLA-DRB51* and *HLA-DRB5*) showed independent evidence of positive selection prior to admixture, based on extended haplotype homozygosity in African populations. A number

of traits related to inflammation, blood metabolites, and both the innate and adaptive immune system showed evidence of admixture-enabled polygenic selection in Latin American populations.

4.1.3 Conclusions

The results reported here, considered together with the ubiquity of admixture in human evolution, suggest that admixture serves as a fundamental mechanism that drives rapid adaptive evolution in human populations.

4.2 Background

Admixture is increasingly recognized as a ubiquitous feature of human evolution [27]. Recent studies on ancient DNA have underscored the extent to which human evolution has been characterized by recurrent episodes of population isolation and divergence followed by convergence and admixture. In this study, we considered the implications of admixture for human adaptive evolution [26]. We hypothesized that admixture is a critical mechanism that enables rapid adaptive evolution in human populations, and we tested this hypothesis via the analysis of admixed genome sequences from four Latin American populations: Colombia, Mexico, Peru, and Puerto Rico. We refer to the process whereby the presence of distinct ancestry-specific haplotypes on a shared population genomic background facilitates adaptive evolution as 'admixture-enabled selection'.

The conquest and colonization of the Americas represents a major upheaval in the global migration of our species and is one of the most abrupt and massive admixture events

known to have occurred in human evolution [3, 4]. The ancestral source populations – from Africa, Europe, and the Americas – that admixed to form modern Latin American populations evolved separately for tens-of-thousands of years before coming together over the last 500 years. This 500-year time frame, corresponding to approximately 20 generations, amounts to less than 1% of the time that has elapsed since modern humans first emerged from Africa [1, 2]. Considered together, these facts point to admixed Latin American populations as an ideal system to study the effects of admixture on rapid adaptive evolution in humans [17].

A number of previous studies have considered the possibility of admixture-enabled selection in the Americas, yielding conflicting results. On the one hand, independent studies have turned up evidence for admixture-enabled selection at the major histocompatibility complex (MHC) locus in Puerto Rico [53], Colombia [56], and Mexico [58], and another study found evidence for admixture-enabled selection on immune system signaling in African-Americans, particularly as it relates to influenza and malaria response [59]. Together, these studies highlighted the importance of the immune system as a target for admixture-enabled selection among a diverse group of admixed American populations. However, a follow up study on a different cohort of African-Americans found no evidence for admixture-enabled selection in the Americas [60]. The latter study concluded that the observed differences in local ancestry reported by previous studies, which were taken as evidence for selection, could have occurred by chance alone given the large number of hypotheses that were tested (i.e. the number of loci analyzed across the genome). This work underscored the importance of controlling for multiple hypothesis testing when investigating the possibility of admixture-enabled selection in the Americas.

We attempted to resolve this conundrum by performing integrated analyses that combine information from (1) single loci across multiple populations and (2) multiple loci that encode polygenic traits. We also used admixture simulation, along with additional lines of evidence from haplotype-based selection scans, to increase the stringency of, and confidence in, our screen for admixture-enabled selection. This combined evidence approach has proven to be effective for the discovery of admixture-enabled selection among diverse African populations [49, 87]. We found evidence for admixture-enabled selection at the MHC locus across multiple Latin American populations, consistent with previous results, and our polygenic screen uncovered novel evidence for adaptive evolution on a number of inflammation, blood, and immune related traits.

4.3 Results

4.3.1 Genetic ancestry and admixture in Latin America

We inferred patterns of genetic ancestry and admixture for four Latin American (LA) populations characterized as part of the 1000 Genomes Project: Colombia (*n*=94), Mexico (*n*=64), Peru (*n*=85), and Puerto Rico (*n*=104) (Figure 10). Genome-wide continental ancestry fractions were inferred using the program ADMIXTURE [73], and local (haplotype-specific) ancestry was inferred using the program RFMix [51]. The results from both programs are highly concordant (Figure 16). As expected [65, 85, 88, 89], the four LA populations show genetic ancestry contributions from African, European, and Native American source populations, and they are distinguished by the relative proportions of each ancestry. Overall, these populations show primarily European ancestry followed by Native American and African components. Puerto Rico has the highest

European ancestry, whereas Peru shows the highest Native American ancestry. Mexico shows relatively even levels of Native American and European ancestry, while Colombia shows the highest levels of three-way admixture. Individual genomes vary greatly with respect to the genome-wide patterns of local ancestry, i.e. the chromosomal locations of ancestry-specific haplotypes (Figure 17). If the process of admixture is largely neutral, then we expect ancestry-specific haplotypes to be randomly distributed throughout the genome in proportions corresponding to the genome-wide ancestry fractions.

4.3.2 Ancestry enrichment and admixture-enabled selection

For each of the four LA populations, local ancestry patterns were used to search for specific loci that show contributions from one of the three ancestral source populations which are greater than can be expected based on the genome-wide ancestry proportions for the entire population (Figure 18). The ancestry enrichment metric that we use for this screen (z_{anc}) is expressed as the number of standard deviations above or below the genome-wide ancestry fraction. Previous studies have used this general approach to look for evidence of admixture-enabled selection at individual genes within specific populations, yielding mixed results [53, 56, 58-60]. For this study, we have added two new dimensions to this general approach in an effort to simultaneously increase the confidence for admixture-enabled selection inferences and to broaden the functional scope of previous studies. To achieve these ends, we searched for (1) concordant signals of ancestry enrichment for single genes (loci) across multiple populations, and (2) concordant signals of ancestry enrichment across multiple genes that function together to encode polygenic phenotypes. The first approach can be considered to increase specificity, whereas the

second approach increases sensitivity. Loci that showed evidence for ancestry enrichment using this combined approach were interrogated for signals of positive selection using the integrated haplotype score (iHS) [55] to further narrow the list of potential targets of admixture-enabled selection.

4.3.3 Single gene admixture-enabled selection

Gene-specific ancestry enrichment values (z_{anc}) were computed for each of the three continental ancestry components within each of the four admixed LA populations analyzed here. We then integrated gene-specific z_{anc} values across the four LA populations using a Fisher combined score (F_{CS}). The strongest signals of single gene ancestry enrichment were seen for African ancestry at the major histocompatibility complex (MHC) locus on the short arm of chromosome 6 (Figure 11A). Three out of the four LA populations show relatively high and constant ancestry enrichment across this locus, with the highest levels of enrichment seen for Mexico and Colombia (Figure 11B). This signal is robust to control for multiple statistical tests using the Benjamini–Hochberg false discovery rate (FDR).

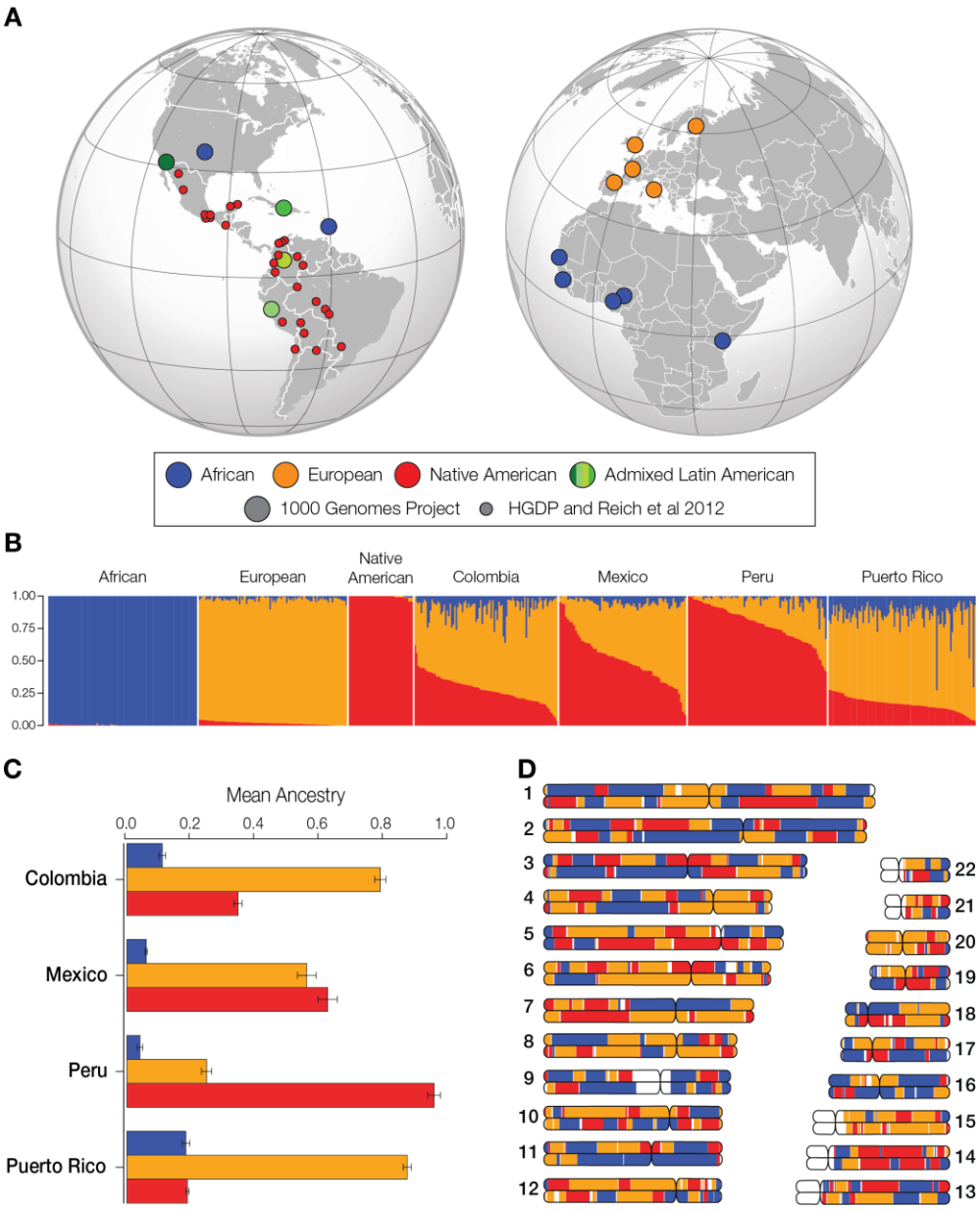

Figure 10. Genetic ancestry and admixture in Latin America.

(A) The global locations of the four LA populations analyzed here (green) are shown along with the locations of the African (blue), European (orange), and Native American (red) reference populations. The sources of the genomic data are indicated in the key. (B) ADMIXTURE plot showing the three-way continental ancestry components for individuals from the four LA populations – Colombia, Mexico, Peru, and Puerto Rico – compared to global reference populations. (C) The mean ($\pm se$) continental ancestry fractions for the four LA populations. (D) Chromosome painting showing the genomic locations of ancestry-specific haplotypes for an admixed LA genome.

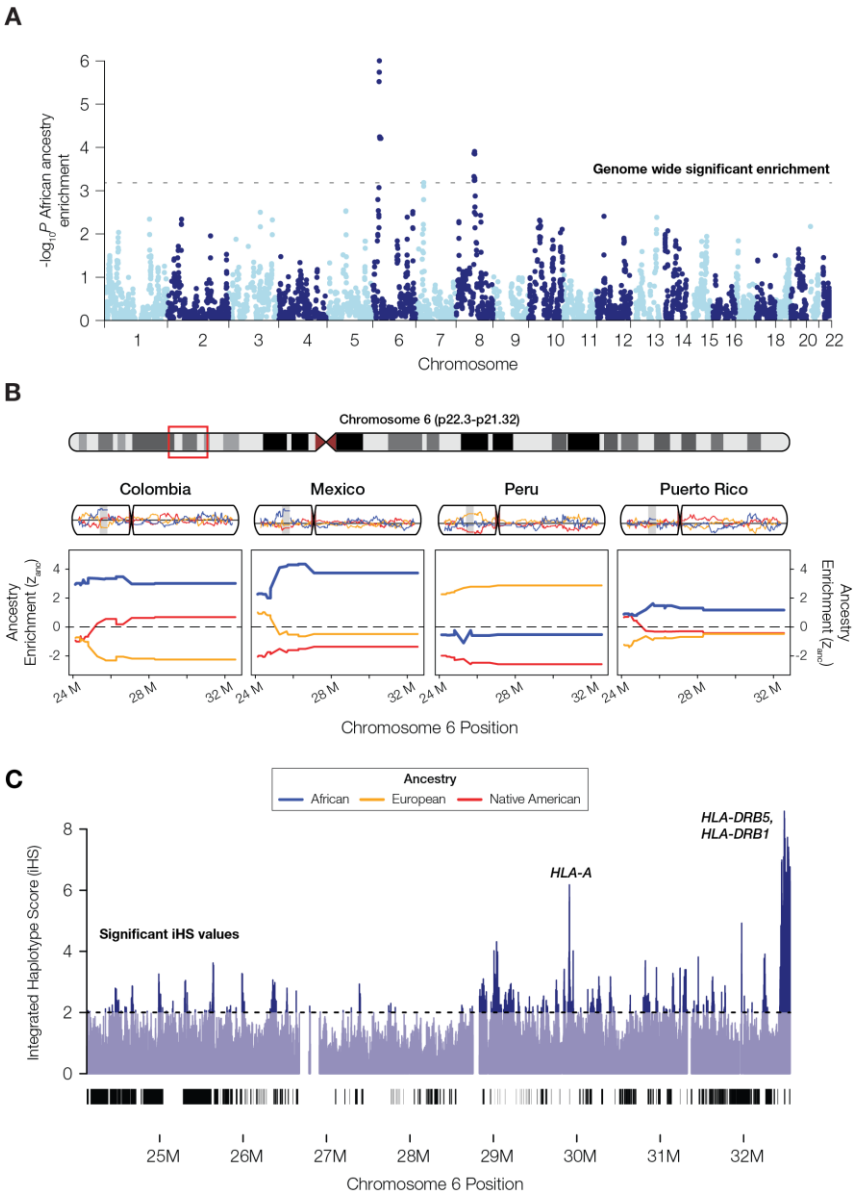

Figure 11. African ancestry enrichment at the major histocompatibility complex (MHC) locus.

(A) Manhattan plot showing the statistical significance of African ancestry enrichment across the genome. (B) Haplotype on chromosome 6 with significant African ancestry enrichment for three of the four LA populations: Colombia, Mexico, and Puerto Rico. This region corresponds to the largest peak of African ancestry enrichment on chromosome 6 seen in panel A. Population-specific African (blue), European (orange), and Native American (red) ancestry enrichment values (z_{anc}) are shown for chromosome 6 and the MHC locus. (C) Integrated haplotype score (iHS) values for African continental population from the 1KGP are shown for the MHC locus; peaks correspond to putative positively selected human leukocyte antigen (HLA) genes.

We simulated random admixture across the four LA populations, parameterized by their genome-wide ancestry proportions, to further assess the probability that this signal could be generated by chance alone (i.e. by genetic drift). Based on this simulation, the observed levels of cross-population African ancestry enrichment at the MHC locus are highly unlikely to have occurred by chance ($P < 5 \times 10^{-5}$), whereas the observed patterns of European and Native American ancestry enrichment are consistent with the range of expected levels generated by the random admixture simulation (Figure 19). Results of the admixture simulation analysis were also used to demonstrate that the cross-population approach to single locus ancestry-enrichment is sufficiently powered to detect selection at the population sizes analyzed here (Figures 20 and 21). The statistical power of the ancestry enrichment approach used in this study rests on the cross-population comparisons, as the probability of observing the same ancestry enrichment at the same locus across multiple LA populations is diminishingly low.

The MHC locus of chromosome 6 also shows a number of peaks for the iHS metric of positive selection from the African continental population (Figure 11). These peaks rise well above the value of 2, which is taken as a threshold for putative evidence of positive selection [55]. The highest African iHS scores are seen for the human leukocyte antigen (HLA) encoding genes *HLA-A*, *HLA-DRB5*, and *HLA-DRB1* (Figure 12A and 12B). These HLA protein encoding genes make up part of the MHC class I (*HLA-A*) and MHC class II (*HLA-DRB5*, and *HLA-DRB1*) antigen presenting pathways of the adaptive immune system (Figure 12C), consistent with shared selective pressures on immune response in admixed LA populations.

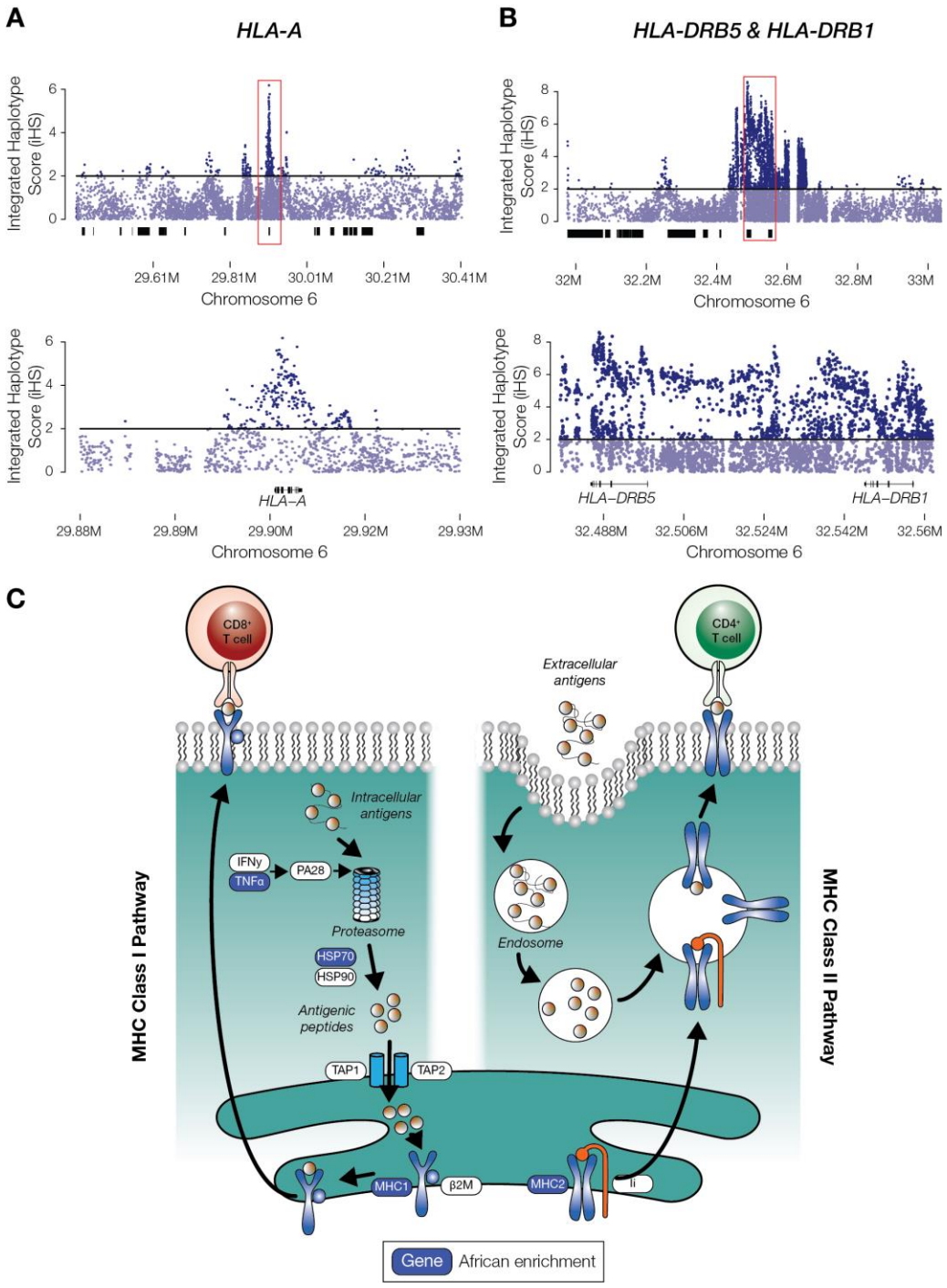

Figure 12. Admixture-enabled selection at human leukocyte antigen (HLA) genes.

Integrated haplotype score (iHS) peaks for the African continental population from the 1KGP are shown for (A) the MHC Class I gene HLA-A and (B) the MHC Class II genes HLA-DRB5 and HLA-DRB1. (C) Illustration of the MHC Class I and MHC Class II antigen presenting pathways, with African enriched genes shown in blue.

We modeled the magnitude of selection pressure that would be needed to generate the observed levels of cross-population African ancestry enrichment at the MHC locus, using a tri-allelic recursive population genetics model that treats ancestry haplotype fractions as allele frequencies (Figure 13). The average selection coefficient value for African MHC haplotypes is s=0.05 (Figure 22), indicating strong selection at this locus over the last several hundred years since the admixed LA populations were formed, consistent with previous work [58].

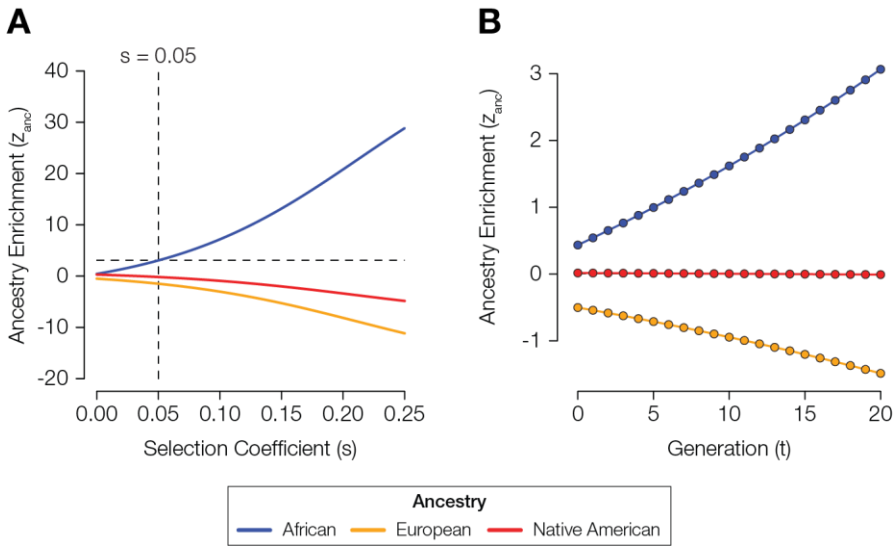

Figure 13. Model of ancestry-enabled selection at the MHC locus of the Colombian population.

(A) Modeled levels of ancestry enrichment and depletion (z_{anc}, y-axis) corresponding to a range of different selection coefficients (s, x-axis): African (blue), European (orange), and Native American (red). The intersection of the observed level of African ancestry enrichment at the MHC locus and the corresponding s-value is indicated with dashed lines. (B) The trajectory of predicted ancestry enrichment and depletion (z_{anc}, y-axis) over time (t generations, x-axis) is shown for the inferred selection coefficient of s=0.05.

4.3.4 Polygenic admixture-enabled selection

For each of the three continental ancestry components, we combined gene-specific ancestry enrichment values (z_{anc}), for genes that function together to encode polygenic phenotypes, via the polygenic ancestry enrichment score (PAE) (Figure 14A). Observed PAE values were compared to expected values generated by randomly permuting size-matched gene sets to search for functions (traits) that show evidence of admixture-enabled selection (Figure 23). As with the single locus approach, we narrowed our list of targets to traits that showed evidence of polygenic admixture enrichment across multiple LA populations. This approach yielded evidence of statistically significant enrichment and depletion, across multiple ancestries, for a number of inflammation, blood, and immune related traits (Figure 14B). Inflammation related phenotypes that show polygenic ancestry enrichment include a variety of skin conditions and rheumatoid arthritis. A number of different blood metabolite pathways show evidence for primarily European and Native American ancestry enrichment, while both the adaptive and innate components of the immune system show evidence of admixture-enabled selection.

Several interconnected pathways of the innate immune system – the RIG-I-like receptor signaling pathway, the Toll-like receptor signaling pathway, and the cytosolic DNA-sensing pathway – all show evidence of Native American ancestry enrichment (Figure 15). All three of these pathways are involved in rapid, first line immune response to a variety of RNA and DNA viruses as well as bacterial pathogens. Genes from these pathways that show evidence of Native American ancestry enrichment encode a number of distinct interferon, interleukin, and cytokine proteins.

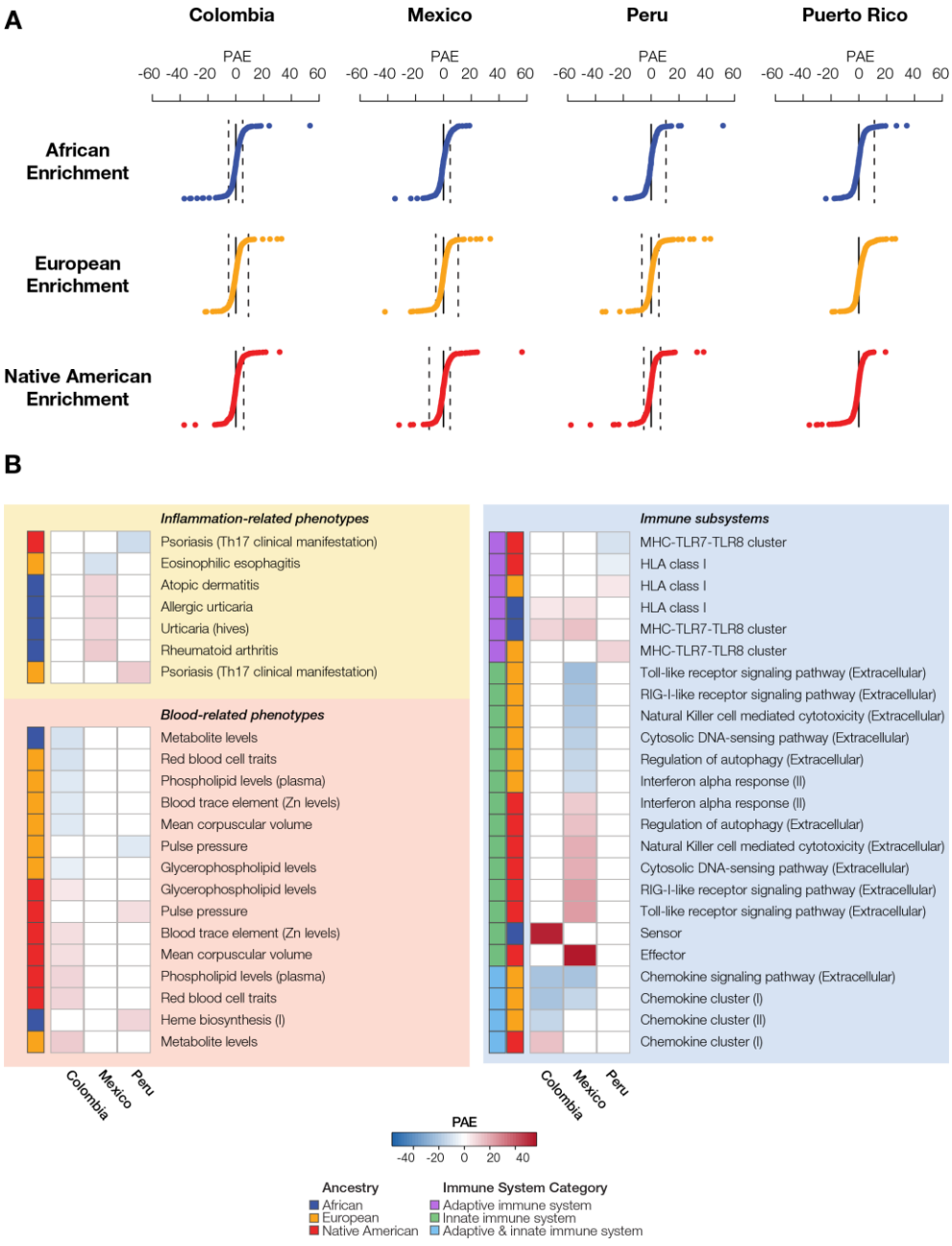

Figure 14. Polygenic ancestry enrichment (*PAE*) and admixture-enabled selection.

(A) Distributions of the PAE test statistic are shown for each of the three ancestry components – African (blue), European (orange), and Native American (red) – across the four LA populations. Points beyond the dashed lines correspond to polygenic traits with statistically significant PAE values, after correction for multiple tests. (B) Polygenic traits that show evidence of PAE in multiple LA populations. PAE values are color coded as shown in the key, and the ancestry components are indicated for each trait. Immune system traits are divided into adaptive (purple), innate (green), or both (blue).

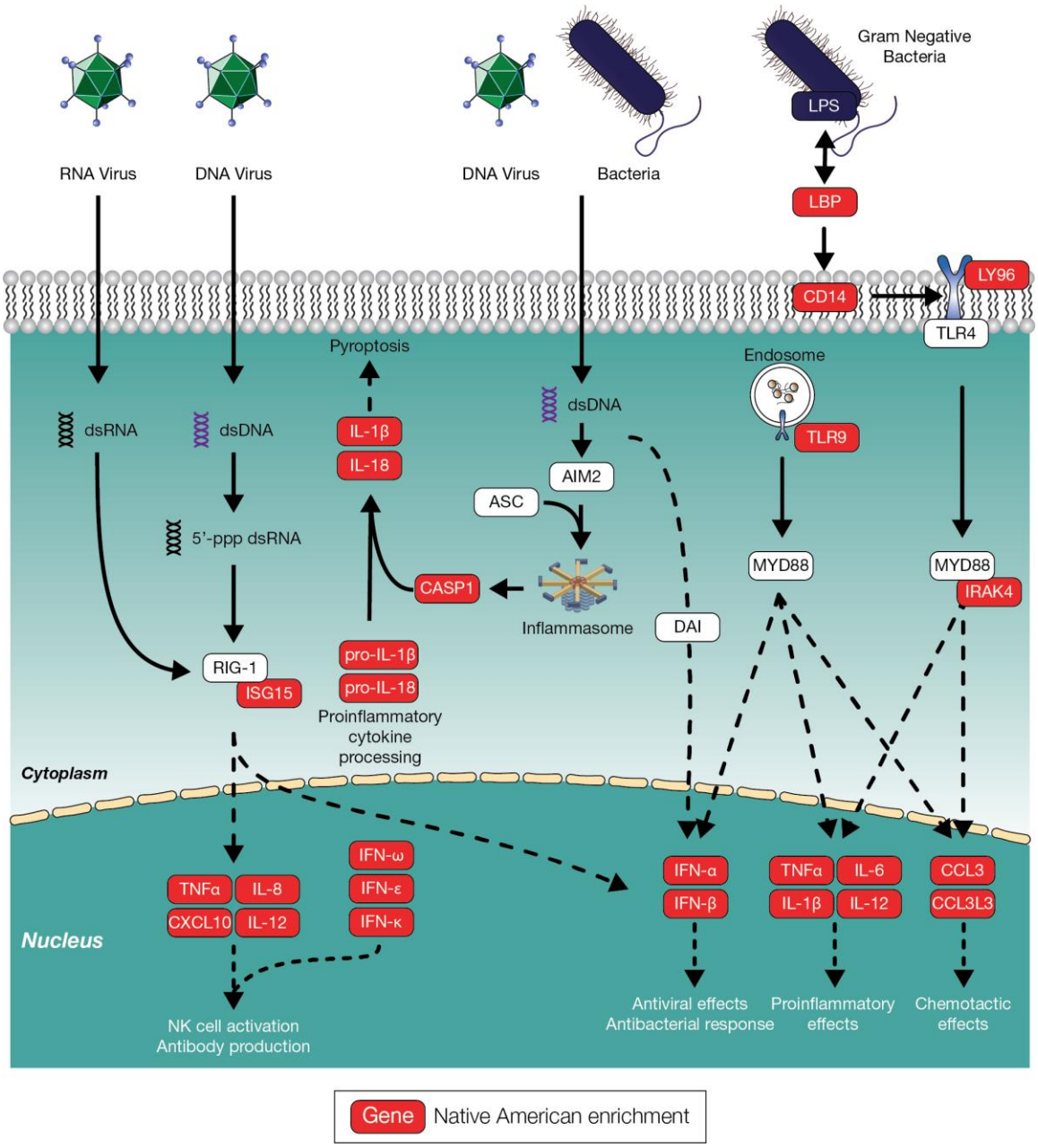

Figure 15. Innate immune system pathways showing Native American ancestry enrichment.

Illustration of three interconnected pathways from the innate immune system – the RIG-I-like receptor signaling pathway, the Toll-like receptor signaling pathway, and the cytosolic DNA-sensing pathway – highlighting genes (proteins) that show Native American ancestry enrichment.

4.4 Discussion

4.4.1 Rapid adaptive evolution in humans

Human adaptive evolution is often considered to be a slow process, which is limited by relatively low effective population sizes and long generation times [32, 90, 91]. The rate of human adaptive evolution is further constrained by the introduction of new mutations [92]. Initially, positive selection acts very slowly to gradually increase the frequency of newly introduced beneficial mutations, which by definition are found at low population frequencies. The process of admixture, whereby previously diverged populations converge, brings together haplotypes that have not previously existed on the same population genomic background [50]. In so doing, it can provide raw material for rapid adaptive evolution in the form of novel variants that are introduced at intermediate frequencies, many of which may have pre-evolved adaptive utility [17].

4.4.2 Admixture and rapid adaptive evolution

Our results suggest that admixture can enable extremely rapid adaptive evolution in human populations. In the case of the LA populations studied here, we found evidence of adaptive evolution within the last 500 years (or ~20 generations) since the conquest and colonization of the Americas began [3, 4]. We propose that, given the ubiquity of admixture among previously diverged populations [26, 27], it should be considered as a fundamental mechanism for the acceleration of human evolution.

The haplotypes that show evidence of ancestry enrichment in our study evolved separately for tens-of-thousands of years in the ancestral source populations – African,

European, and Native American – that mixed to form modern, cosmopolitan LA populations. Many of these haplotypes are likely to contain variants, or combinations of variants, that provided a selective advantage in their ancestral environments [34]. These adaptive variants would have increased in frequency over long periods of time and then later provided source material for rapid adaptation of admixed populations, depending on their utility in the New World environment. Variants that reached high frequency in ancestral source populations via genetic drift could also serve as targets for positive selection in light of the distinct environments and selection pressures faced by modern admixed populations. In either case, admixture-enabled selection can be taken as a special case of selection on standing variation, or soft selective sweeps, underscoring its ability to support rapid adaptation in the face of novel selective pressures [93, 94].

4.4.3 Single locus versus polygenic selection

Our initial analysis of individual LA populations turned up numerous instances of apparent ancestry enrichment genome-wide, including enrichment for all three ancestry components in each of the four populations studied here. However, when ancestry enrichment signals were combined across all four populations, only a handful of significant results remained after correcting for multiple tests. Finally, when random admixture was simulated, only two peaks of African ancestry enrichment were found to be shared among populations at levels greater than expected by chance (Figures 11 and 19). These findings support the conservative nature of our combined evidence approach to using cross-population ancestry enrichment as a criterion for inferring admixture-enabled selection, and also reflect the fact that selection needs to be extremely strong to be detected at single loci. This is especially true given the relatively short period of time that has elapsed since

modern LA populations were formed via admixture of ancestral source populations. The results of our population genetic model support this notion, showing an average selection coefficient value of s=0.05 for African haplotypes at the MHC locus.

A number of recent studies have underscored the ubiquity of polygenic selection on complex traits that are encoded by multiple genes, emphasizing the fact that weaker selection dispersed across multiple loci may be a more common mode of adaptive evolution than strong single locus selection [95-98]. The results of our polygenic ancestry enrichment analysis are consistent with these findings, as the polygenic approach yielded signals of admixture-enabled selection for numerous traits across different ancestry components and populations. Thus, the polygenic ancestry enrichment that we employed to infer admixture-enabled selection is both more biologically realistic and better powered compared to the single locus approach.

4.5 Conclusions

We report abundant evidence for admixture-enabled selection within and between Latin American populations that were formed by admixture among diverse African, European, and Native American source populations within the last 500 years. The MHC locus shows evidence of particularly strong admixture-enabled selection for several *HLA* genes, all of which appear to contain pre-adapted variants that were selected prior to admixture in the Americas. In addition, a number of related immune system, inflammation and blood metabolite traits were found to evolve via polygenic admixture-enabled selection.

Over the last several years, it has become increasingly apparent that admixture is a ubiquitous feature of human evolution. Considering the results of our study together with the prevalence of admixture leads us to conclude that admixture-enabled selection has been a fundamental mechanism for driving rapid adaptive evolution in human populations.

4.6 Methods

4.6.1 Genomic data

Whole genome sequence data for four admixed LA populations – Colombia, Mexico, Peru, and Puerto Rico – were taken from the Phase 3 data release of the 1000 Genomes Project (1KGP) [70, 99]. Whole genome sequence data and whole genome genotypes for proxy ancestral reference populations from Africa, Asia, Europe, and the Americas were taken from multiple sources, including the 1KGP, the Human Genome Diversity Project (HGDP) [64] and a previous study on Native American genetic ancestry [100] (Table 2). Whole genome sequence and whole genome genotype data were harmonized using the program PLINK [71], keeping only those sites common to all datasets and correcting SNP strand orientations as needed. A genotyping filter of 95% calls was applied to all populations.

4.6.2 Global and local ancestry inference

Global continental ancestry estimates for each individual from the four LA populations were inferred using the program ADMIXTURE [73]. The harmonized SNP set was pruned using PLINK [71] with window size of 50bp, a step size of 10bp, and a linkage disequilibrium (LD) threshold of $r^2 > 0.1$, and ADMIXTURE was run with *K*=4

corresponding to African, European, Asian, and Native American ancestry components. Local continental ancestry estimates for each individual from the four LA populations were inferred using a modified version of the program RFMix [51] as previously described [101]. The chromosomal locations of ancestry-specific haplotypes were visualized with the program Tagore [102]. The complete harmonized SNP set was phased using the program SHAPEIT [72], and RFMix was run to assign African, European, or Native American ancestry to individual haplotypes from the LA populations.

4.6.3 Single locus ancestry enrichment

Single gene (locus) ancestry enrichment (z_{anc}) values were calculated for all three continental ancestry components (African, European, and Native American) across all four LA populations. Genomic locations of NCBI RefSeq gene models were taken from the UCSC Genome Brower (hg19 build) [103], and gene locations were mapped to the ancestry-specific haplotypes characterized using RFMix for each individual genome. For each gene, population-specific three-way ancestry fractions (f_{anc}) were computed as the number of ancestry-specific haplotypes (h_{anc}), divided by the total number of ancestry-assigned haplotypes for that gene (h_{tot}): $f_{anc} = h_{anc}/h_{tot}$. Ancestry enrichment analysis was limited to genes that had h_{tot} values within one standard deviation of the genome-wide average for any population. Distributions of gene-specific ancestry fractions (f_{anc}) for each population were used to calculate population-specific genome-wide average (μ_{anc}) and standard deviation (σ_{anc}) ancestry fractions. Then, for any given gene in any given population, ancestry enrichment (z_{anc}) was calculated as the number of standard deviations above (or below) the genome-wide ancestry average: $z_{anc} =$

$(f_{anc} - \mu_{anc})/\sigma_{anc}$, with gene-specific ancestry enrichment *P*-values computed using the z distribution. A Fisher's combined score (F_{CS}) was used to combine gene-specific ancestry enrichment *P*-values across the four LA populations as: $F_{CS} = -2\sum_{i=1}^{4} \ln(P_i)$. The statistical significance of F_{CS} was computed using the χ^2 distribution with 8 ($2k$) degrees of freedom. Correction for multiple F_{CS} tests was performed using the Benjamini-Hochberg False Discovery Rate (FDR), with a significance threshold of $q < 0.05$ [104].

4.6.4 Admixture simulation

Three-way admixed individuals were randomly simulated for each LA population – Colombia, Mexico, Peru, and Puerto Rico – and used to calculate expected levels of ancestry enrichment z_{anc} as described in the previous section. Expected levels of z_{anc} were combined across the four LA populations to yield expected Fisher's combined scores (F_{CS}) and their associated *P*-values as described in the previous section. Admixed populations were simulated as collections of genes (i.e. ancestry-specific haplotypes) randomly drawn from the genome-wide ancestry distributions for each LA population. Sized matched admixed populations were simulated for each LA population and combined to generate expected (F_{CS}) and their associated *P*-values, and admixture simulation was also conducted across a range of population sizes (n=10 to 10,000) to evaluate the power of the combined evidence cross-population approach used to detect ancestry-enabled selection.

4.6.5 Polygenic ancestry enrichment

Polygenic ancestry enrichment values (PAE) were computed by combining single locus ancestry enrichment values (z_{anc}) across genes that function together to encode

polygenic traits. Gene sets for polygenic traits were curated from a number of literature and database sources to represent a wide array of phenotypes (Table 3). All gene sets were LD pruned with a threshold of $r^2 > 0.1$ using PLINK. Additional details on the curation of polygenic trait gene sets can be found in section *4.7.2.1 Polygenic trait gene set curation*. For any trait trait-specific gene set, in any population, PAE was calculated by summing the gene-specific z_{anc} values for all of the genes in the trait set: $PAE = \sum_1^n z_{anc}$, where n is the number of genes in the set. Since z_{anc} values can be positive or negative, depending on over- or under-represented ancestry, values of PAE are expected to be randomly distributed around 0. The statistical significance levels of observed PAE values were calculated via comparison against distributions of expected PAE values calculated from 10,000 random permutations of size matched gene sets (Figure 23). Observed values (PAE_{obs}) were compared against the mean (μ_{PAE}) and standard deviation (σ_{PAE}) of the expected PAE values to compute the statistical significance for each trait: $z_{PAE} = (PAE_{obs} - \mu_{PAE})/\sigma_{PAE}$, with P-values computed using the z distribution Correction for multiple tests was performed using the Benjamini-Hochberg False Discovery Rate (FDR), with a significance threshold of $q < 0.05$.

4.6.6 Integrated Haplotype Scores (iHS)

Integrated Haplotype Scores (iHS) [55] were calculated for European and African continental populations from the 1KGP using the software selscan (version 1.1.0a) [105]. |iHS| scores were overlaid on genes with evidence of ancestry enrichment to scan for concurrent signals of selection.

4.6.7 Modeling admixture-enabled selection

Admixture-enabled selection was modeled for the African enriched chromosome 6 MHC haplotype using a standard recursive population genetics model for positive selection [106]. Three allelic states were used for the selection model, each of which corresponds to a specific ancestry component: African, European, or Native American. Population-specific models were initialized with allele (ancestry) frequencies based on the genome-wide background ancestry fractions and run across a range of selection coefficient (s) values to determine the values of s that correspond to the observed African ancestry enrichment levels. This allowed us to compute a positive selection coefficient corresponding to the strength of African ancestry selection at the MHC locus for each population. Additional details of this model can be found in section *4.7.1 Modeling admixture-enabled selection at single loci* and Figure 22.

4.7 Supplementary Methods and Results

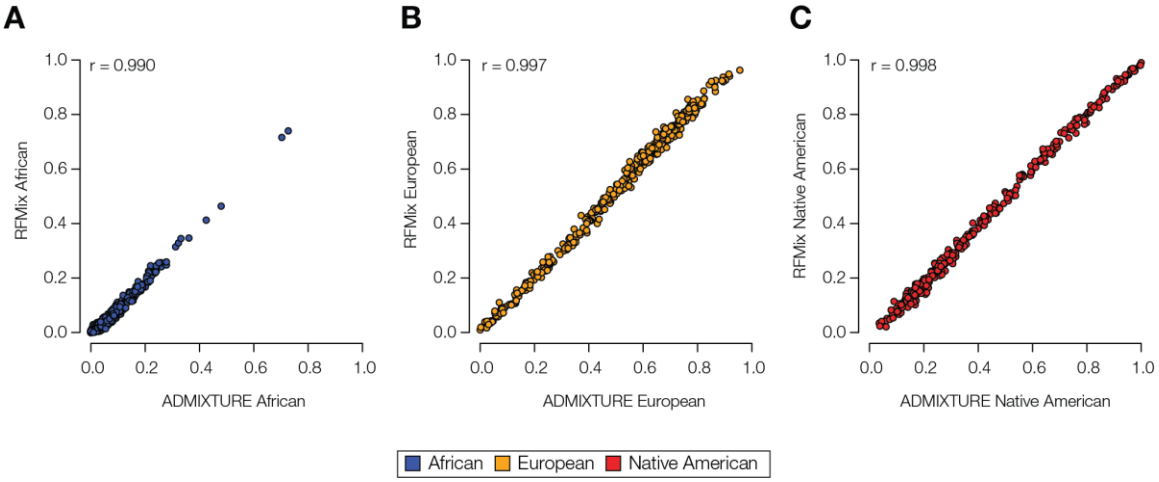

Figure 16. Correspondence between continental ancestry estimates for LA populations generated by ADMIXTURE (x-axis) and RFMix (y-axis).

Ancestry estimates generated for individuals from the four LA populations studied here – Colombia, Mexico, Peru, and Puerto Rico – are shown separately for (A) African, (B) European, and (C) Native American ancestry. Correlation between ancestry estimates generated using the two programs are measured by Pearson's correlation coefficient (r).

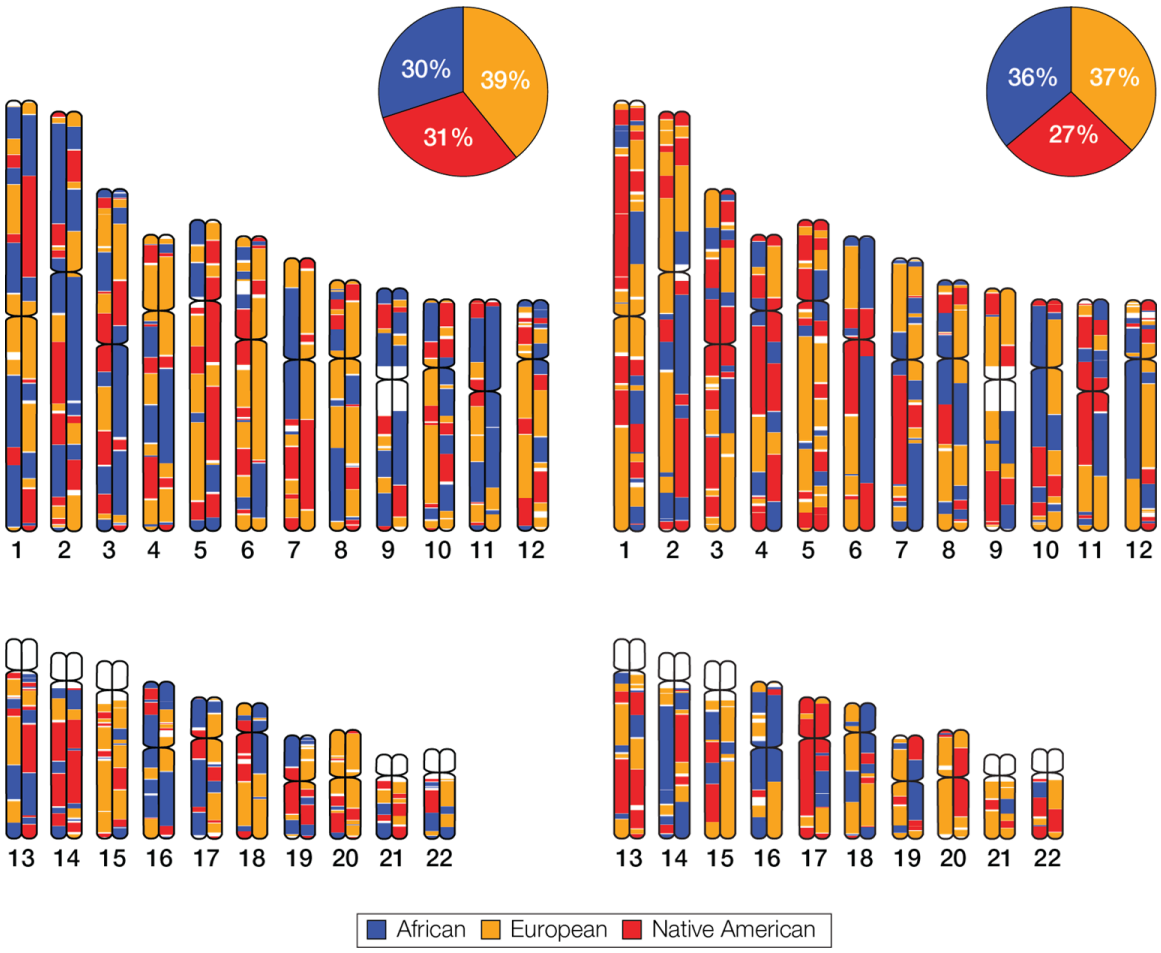

Figure 17. Comparison of global versus local continental ancestry inference for two admixed individuals.

Continental ancestry estimates – for African (blue), European (orange), and Native American (red) ancestry components – were inferred using global (ADMIXTURE) and local (RFMix) ancestry methods. Global ancestry estimates are shown as pie charts, and local ancestry estimates (i.e. ancestry-specific haplotype assignments) are shown as chromosome ideograms. These two individuals have very similar levels of global ancestry but highly distinct local ancestry patterns.

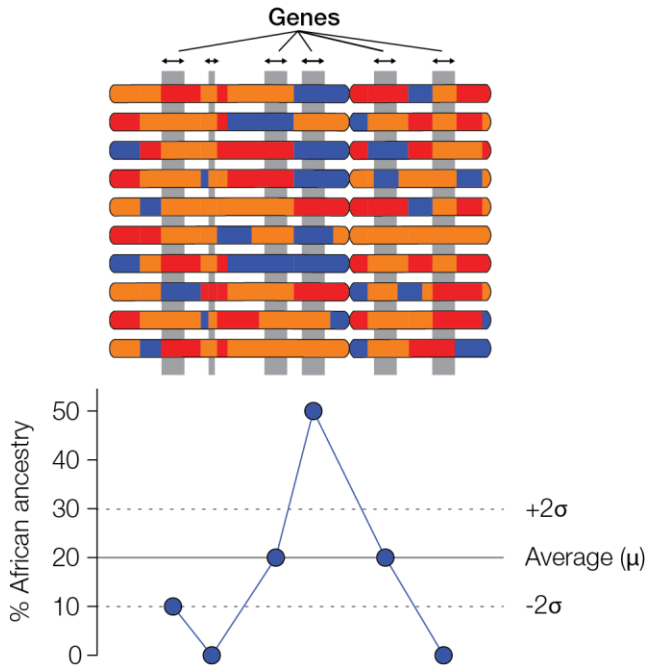

Figure 18. Scheme of the ancestry enrichment method used for this study.

Ten haploid chromosomes from a population are shown with corresponding regions aligned and ancestry-specific haplotypes indicated (as in Figure S2). Gene-specific ancestry enrichment (z_{anc}) is expressed as the number of standard deviations above or below the genome-wide ancestry fraction for the population.

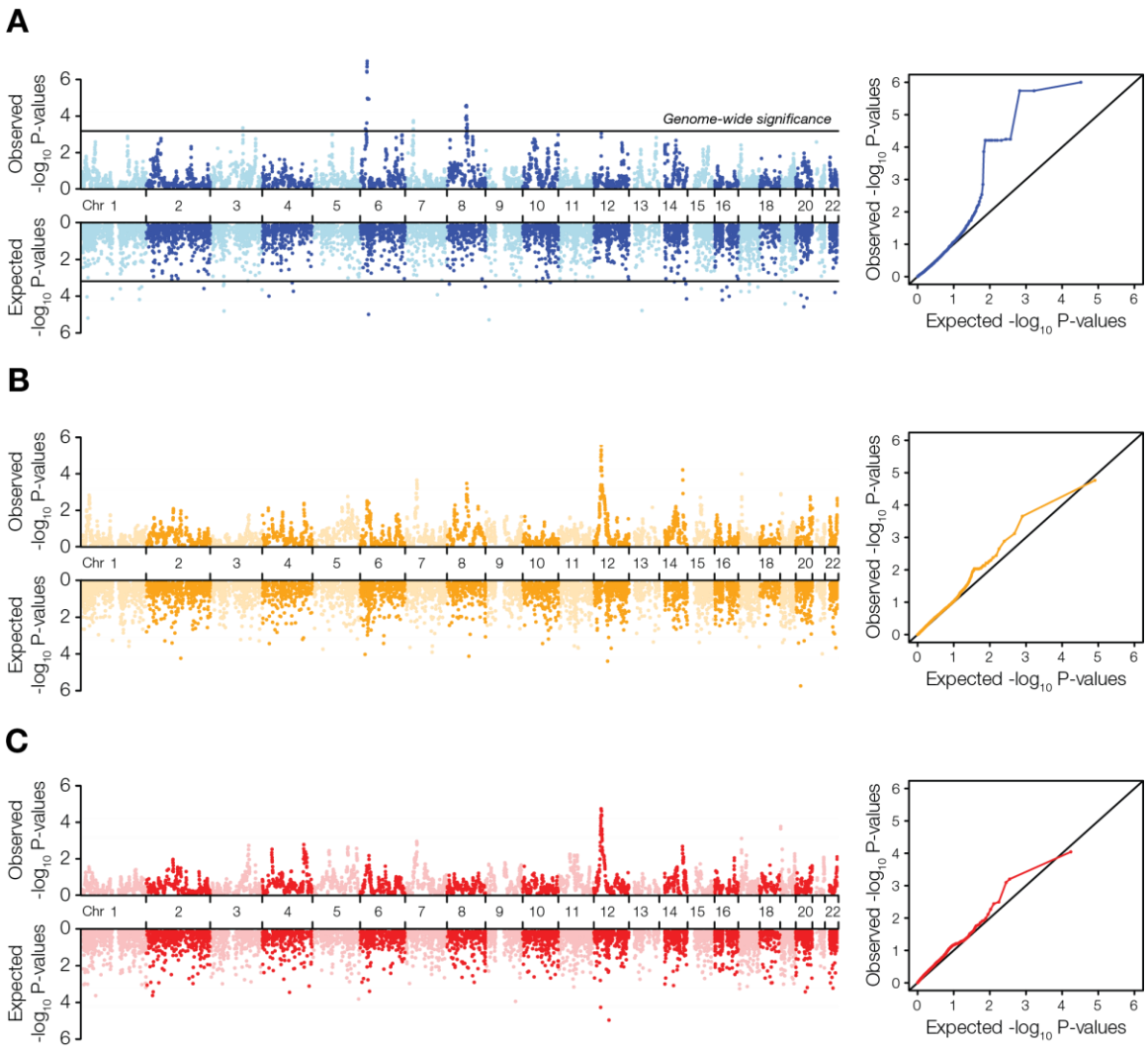

Figure 19. Observed versus expected ancestry enrichment across the four LA populations studied here.

Manhattan plots and QQ plots show the genome-wide distributions of observed versus expected ancestry enrichment $-log_{10}$ P values for African (A), European (B), and Native American (C) ancestry components. P-values correspond to the combined cross-population ancestry enrichment values (F_{CS}), calculated as described in the manuscript. Expected values were generated based on size matched simulated admixed populations, as described in the Methods section. On the African ancestry Manhattan plot, lines indicate genome-wide statistically significant cross-population ancestry enrichment (FDR $q < 0.05$). There were no genome-wide significant cross-population ancestry enrichment peaks for the European and Native American ancestry components. Ancestry-specific QQ plots also illustrate the deviation of observed versus expected ancestry enrichment for the African component, and the similarity of observed versus expected ancestry enrichment for the European and Native American components.

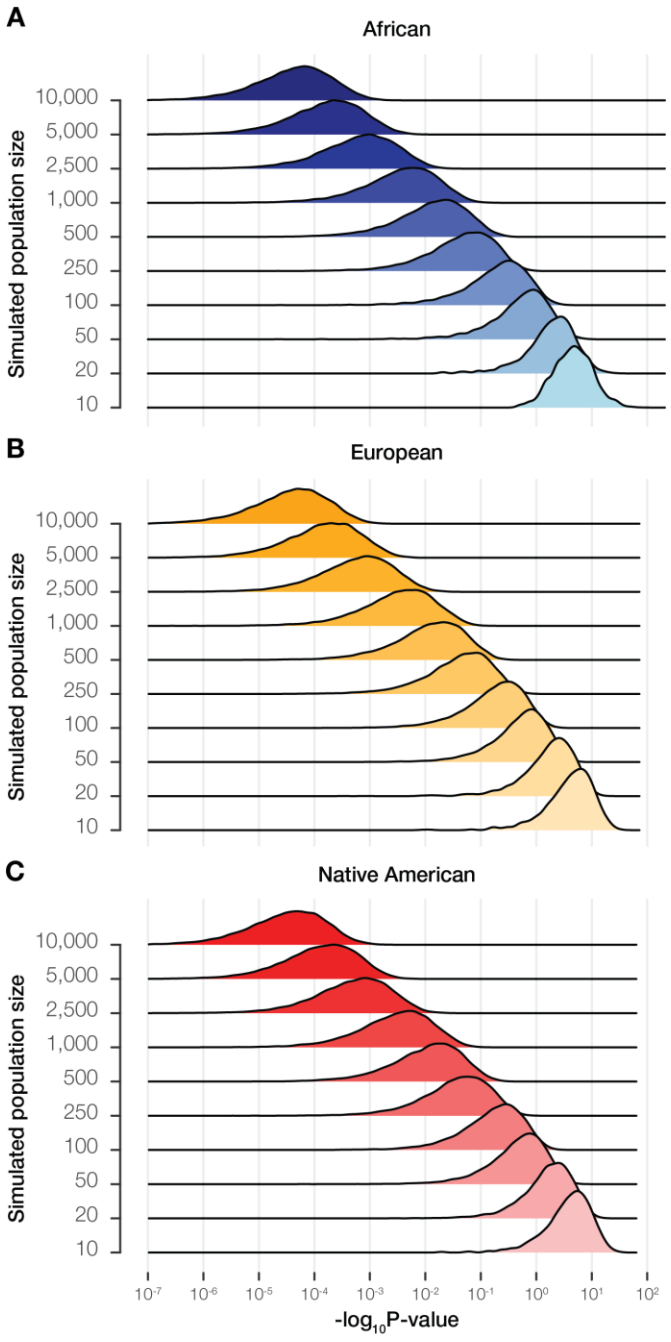

Figure 20. Ancestry enrichment power analysis.

Four admixed LA populations were simulated across a range of population sizes (see Methods). Distributions of $-log_{10}$ P values corresponding to combined cross-population ancestry enrichment values (F_{CS}) are shown across the range of simulated population sizes for each ancestry component: African (A), European (B), and Native American (C). This analysis was used as an additional, non-parametric method to compute the probability of observing ancestry-specific F_{CS} by chance alone, at different population sizes.

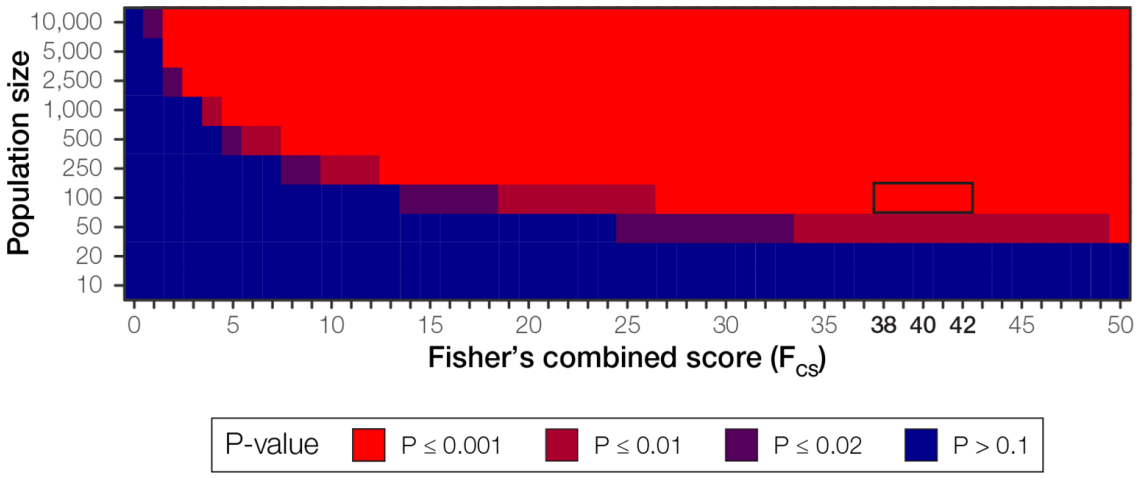

Figure 21. African ancestry enrichment power analysis.

Four admixed LA populations were simulated across a range of population sizes (see Methods). Cross-population African enrichment values (F_{CS}), and their corresponding P values, were computed across the range of population sizes (see Figure S5). The plot shows the relationship of F_{CS} and P-values for different population sizes, i.e. the relationship between population size and the power of the cross-population ancestry enrichment metric. For African ancestry, a statistically significant FDR $q < 0.05$ corresponds to P<0.001 (bright red). Significant P values are more common at higher sample sizes. The observed African F_{CS} values at the MHC locus range from 38-42 (see boxed cells), which are well powered to detect significant ancestry enrichment at population sizes of ~100 individuals.

4.7.1 Modeling admixture-enabled selection at single loci

A recursive triallelic model of positive selection was used to measure the strength of admixture-enabled selection at the chromosome 6 MHC locus in the Colombia, Mexico, and Peru populations. The model treats alleles as the three ancestry fractions: African, European, and Native American. This model is based on the approach to modelling selection used in the Populus software (https://cbs.umn.edu/populus/overview).

Assuming an African ancestry advantage, the relative fitness (w_{ij}) of a locus with ancestries i and j (ancestry genotypes) is calculated as:

$$w_{ij} = \begin{cases} 1, & if\ i = j = African \\ 1 - sh, & if\ i = African\ or\ j = African \\ 1 - s, & if\ i \neq African\ and\ j \neq African \end{cases} \quad (5)$$

where s is the selection coefficient and h is the dominance coefficient. This can be formulated as a fitness matrix:

		Ancestry j		
		African (A)	European (E)	Native American (N)
Ancestry i	African (A)	1	$1-(sh)$	$1-(sh)$
	European (E)	$1-(sh)$	$1-s$	$1-s$
	Native American (N)	$1-(sh)$	$1-s$	$1-s$

For each population, for each ancestry, the ancestry frequency in the next generation ($p_{i,t+1}$) was calculated as:

$$p_{i,t+1} = \frac{p_{i,t} w_i}{\bar{w}} \quad (6)$$

where $p_{i,t}$ is the ancestry frequency (i) in the current generation (t), w_i is the marginal fitness of the ancestry, and $\bar{w}$ is the population mean fitness.

For each row of the fitness matrix, the marginal fitness (w_i) was calculated as:

$$w_i = \sum_j w_{ij} p_j \quad (7)$$

where w_{ij} is the relative fitness of the ancestry genotype and w_{ij} is the frequency of the ancestry in the current generation.

For each generation, the population mean fitness ($\overline{w}$) was calculated as:

$$\overline{w} = \sum_i \sum_j w_{ij} p_i p_j \quad (8)$$

where w_{ij} is the relative fitness of the ancestry genotype, p_i is the frequency of the first ancestry, and p_j is the second of the second ancestry.

For each population, the model was run using s from 0-1.0, in increments of 0.001, and $h = 0.05$, until the final African ancestry frequency at 20 generations was the same as that observed for the chromosome 6 haplotype. The selection coefficient at this convergence was taken to be the strength of selection for the given population for the chromosome 6 African enriched haplotype.

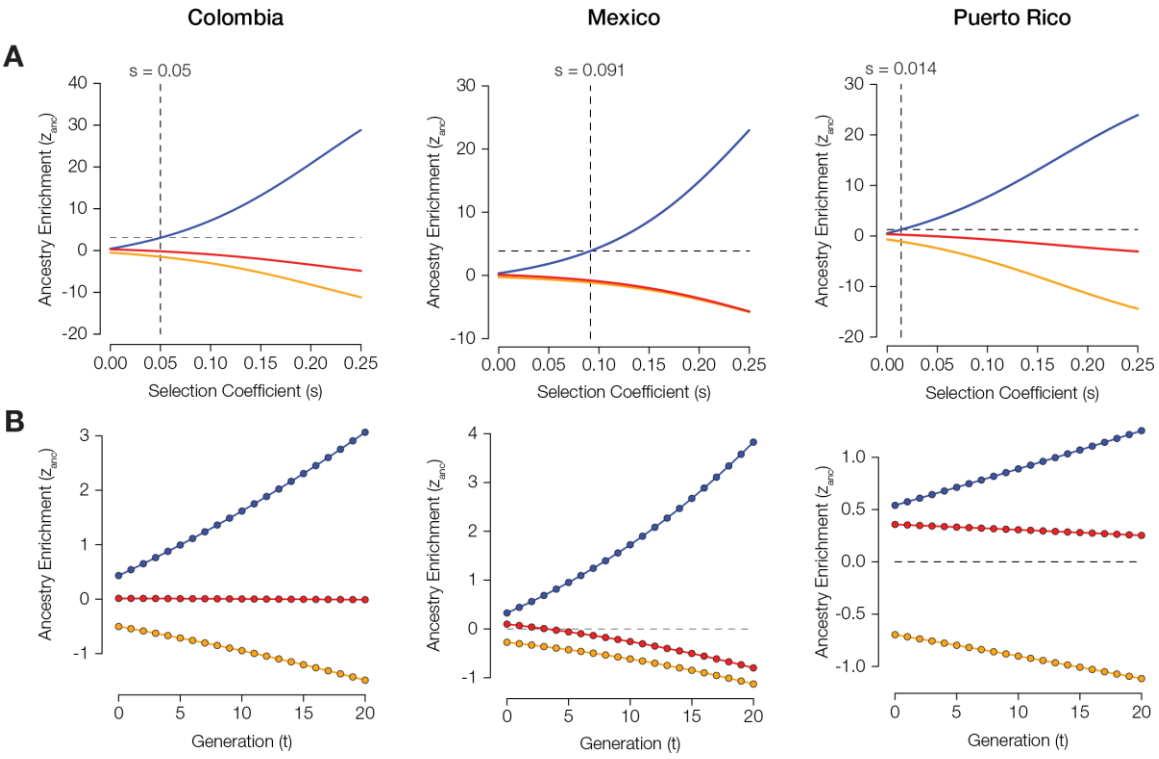

Figure 22. Modeling the strength of admixture-enabled selection at the MHC locus.

The strength of admixture-enabled selection for African haplotypes at the MHC locus was modelled as described on pages 8-9. The three-ancestry population genetic model was run across a range of positive selection coefficient (s) values to identify the strength of selection needed to explain the observed levels of African ancestry enrichment at the MHC locus in the Colombia, Mexico, and Puerto Rico populations. The top panels for each population show the predicted levels of ancestry enrichment (z_{anc}, y-axis) across a range of different selection coefficients (s, x-axis). The intersection of the observed levels of African ancestry enrichment and their corresponding s-values are indicated with dashed lines. The bottom panels show the trajectory of predicted ancestry enrichment and depletion (z_{anc}, y-axis) over time (t generations, x-axis) for each population given the inferred selection coefficient s for each population.

Table 2. Human populations analyzed as part of this study.

Populations are organized into continental groups, for both proxy ancestral reference populations and Latin American populations, and the number of genome (genotype) samples from each population is shown.

Population group	Dataset[1]	Geographical source	*n*	Population group	Dataset[1]	Geographical source	*n*
African	1KGP	Americans of African Ancestry in SW USA	61	Native American	Reich et al	Waunana in Colombia	3
	1KGP	Mende in Sierra Leone	85		Reich et al	Kogi in Colombia	4
	1KGP	African Caribbeans in Barbados	96		Reich et al	Mixtec in Mexico	5
	1KGP	Esan in Nigeria	99		Reich et al	Embera in Colombia	5
	1KGP	Luhya in Webuye, Kenya	99		Reich et al	Guahibo in Colombia	6
	1KGP	Yoruba in Ibadan, Nigeria	108		Reich et al	Ticuna in Brazil	6
	1KGP	Gambian in Western Division in the Gambia	113		Reich et al	Guarani in Paraguay	6
European	1KGP	British in England and Scotland	91		Reich et al	Piapoco in Colombia	7
	1KGP	Finnish in Finland	99		HGDP	Suruí in Brazil	8
	1KGP	Utah Residents (CEPH) with Northern and Western European Ancestry	99		Reich et al	Inga in Colombia	9
	1KGP	Iberian Population in Spain	107		Reich et al	Wayuu in Colombia	11
	1KGP	Toscani in Italia	107		Reich et al	Kaqchikel in Guatemala	13
Latin American	1KGP	Mexican Ancestry in Los Angeles USA	64		HGDP	Pima in Mexico	14
	1KGP	Peruvians from Lima, Peru	85		HGDP	Karitiana in Brazil	14
	1KGP	Colombians from Medellin, Colombia	94		Reich et al	Mixe in Mexico	17
	1KGP	Puerto Ricans from Puerto Rico	104		HGDP	Maya in Mexico	21
					Reich et al	Aymara in Bolivia	23
					Reich et al	Tepehuano in Mexico	25
					Reich et al	Quechua in Peru	40
					Reich et al	Zapotec in Mexico	43

[1]1KGP = 1000 Genomes Project; HGDP = Human Genome Diversity Project; Reich et al [100]

4.7.2 Polygenic ancestry enrichment

4.7.2.1 Polygenic trait gene set curation

Gene sets for polygenic traits were curated from a number of literature and database sources, as shown below, to represent a wide array of phenotypes.

Table 3. Sources of the polygenic trait gene sets analyzed as part of this study.

For each polygenic trait gene set source, the number of phenotypes and reference are shown.

Phenotype source	*n*	Reference
NHGRI-EBI GWAS Catalog	306	[107]
InnateDB	128	[108]
Blood Transcription Modules	209	[109]
Blood Informative Transcripts	9	[110]
Innate Immune System	9	[111]
GIANT	7	[112]
Custom gene sets of interest	10	na

Whenever possible, the gene sets were taken directly from the literature or the database. If gene sets were not directly accessible, SNP-level data was collected and directly mapped to genes to generate the gene set. Trait-specific gene sets from the NHGRI-EBI GWAS Catalog [107] were collected from SNP sets that were mapped using EBI's in-house pipeline; only SNPs that fell within a gene and were implicated at a genome-wide significance level of $P \leq 5\text{x}10^{-8}$ with the phenotype were used to generate the gene sets. Each polygenic trait from the Genetic Investigation of ANthropometric Traits (GIANT) consortium [112] had gene sets mapped according to specifications of the individual paper. All remaining gene sets from the literature were mapped using NCBI's dbSNP database as needed. To control for any linkage between genes, linkage disequilibrium (LD) pruning was performed on each gene set using PLINK. For each set, pruning was performed on all

gene pairs with genic SNP $r^2 > 0.1$; if this was found to be true then only one member of the gene pair was retained for polygenic ancestry enrichment analysis. After LD pruning was complete, the gene sets were filtered based on size so that all phenotypes included in the analysis contained two or more genes.

4.7.2.2 Statistical significance of polygenic ancestry enrichment

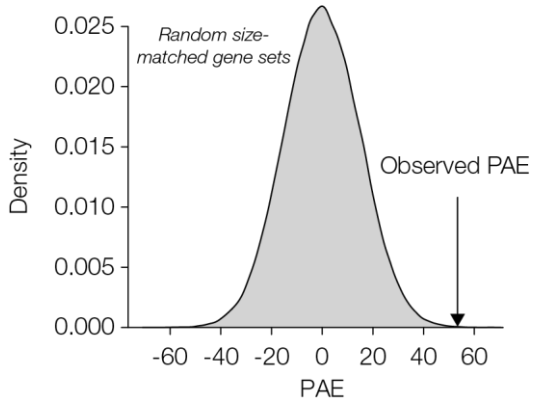

Figure 23. Simulation of random polygenic trait gene sets used to compute statistical significance of polygenic ancestry enrichment (*PAE*).

Size-matched random gene sets are simulated to generate an expected (null) distribution of PAE values.

CHAPTER 5. ASSORTATIVE MATING ON ANCESTRY-VARIANT TRAITS IN ADMIXED LATIN AMERICAN POPULATIONS

5.1 Abstract

Assortative mating is a universal feature of human societies, and individuals from ethnically diverse populations are known to mate assortatively based on similarities in genetic ancestry. However, little is currently known regarding the exact phenotypic cues, or their underlying genetic architecture, which inform ancestry-based assortative mating. We developed a novel approach, using genome-wide analysis of ancestry-specific haplotypes, to evaluate ancestry-based assortative mating on traits whose expression varies among the three continental population groups – African, European, and Native American – that admixed to form modern Latin American populations. Application of this method to genome sequences sampled from Colombia, Mexico, Peru, and Puerto Rico revealed widespread ancestry-based assortative mating. We discovered a number of anthropometric traits (body mass, height, and facial development) and neurological attributes (educational attainment and schizophrenia) that serve as phenotypic cues for ancestry-based assortative mating. Major histocompatibility complex (MHC) loci show population-specific patterns of both assortative and disassortative mating in Latin America. Ancestry-based assortative mating in the populations analyzed here appears to be driven primarily by African ancestry. This study serves as an example of how population genomic analyses can yield novel insights into human behavior.

5.2 Introduction

Mate choice is a fundamental dimension of human behavior with important implications for population genetic structure and evolution [18-20]. It is widely known that humans choose to mate assortatively rather than randomly. That is to say that humans, for the most part, tend to choose mates that are more similar to themselves than can be expected by chance. Historically, assortative mating was based largely on geography, whereby partners were chosen from a limited set of physically proximal individuals [113]. Over millennia, assortative mating within groups of geographically confined individuals contributed to genetic divergence between groups, and the establishment of distinct human populations, such as the major continental population groups recognized today [63, 64, 70].

However, the process of geographic isolation followed by population divergence that characterized human evolution has not been strictly linear. Ongoing human migrations have continuously brought previously isolated populations into contact; when this occurs, the potential exists for once isolated populations to admix, thereby forming novel population groups [114]. Perhaps the most precipitous example of this process occurred in the Americas, starting just over 500 years ago with the arrival of Columbus in the New World [4]. This major historical event quickly led to the co-localization of African, European and Native American populations that had been (mostly) physically isolated for tens of thousands of years [17]. As can be expected, the geographic reunification of these populations was accompanied, to some extent, by genetic admixture and the resulting formation of novel populations. This is particularly true for populations in Latin America, which often show high levels of three-way genetic admixture between continental population groups [56, 65-67, 115].

Nevertheless, modern admixed populations are still very much characterized by non-random assortative mating. Assortative mating in modern populations has been shown to rest on a variety of traits, including physical (stature and pigmentation) and neurological (cognition and personality) attributes. For example, numerous studies have demonstrated an influence of similarities in height and body mass on mate choice [19, 116-118]. In addition, assortative mating has been observed for diverse neurological traits, such as educational attainment, introversion/extroversion and even neurotic tendencies [119-124]. Harder to classify traits related to personal achievement (income and occupational status) and culture (values and political leanings) also impact patterns of assortative mating [119, 125, 126]. Odor is one of the more interesting traits implicated in mate choice, and it has been linked to so-called disassortative (or negative assortative) mating, whereby less similar mates are preferred. Odor-based disassortative mating has been attributed to differences in genes of the major histocompatibility (MHC) locus, which functions in the immune system, based on the idea that combinations of divergent human leukocyte antigen (HLA) alleles provide a selective advantage via elevated host resistance to pathogens [127, 128].

The traits that influence human mate choice are shaped by multiple factors with contributions from genes (G) and the environment (E) along with gene-by-environment (GxE) interactions. Studies that consider both genes and environment have shown different contributions of these factors to assortative mating patterns in human populations. Twin studies were initially used in an effort to tease apart the genetic and environmental contributions to mate choice [129]. Comparison of monozygotic and dizygotic twin pairs provided the first evidence for genetic influences on human mate choice, with 10-30% of

the variance in mate choice explained by genetics compared to 10% shared environment and 60% unique environmental variance [130]. Subsequent twin design studies either did not find any strong genetic effects on patterns of assortative mating [131] or found genetic effects on assortative mating with very different relative contributions of genes versus environmental effects depending on the trait under consideration [132]. A more recent study leveraged genome-wide association study (GWAS) variants that influence height to show even more clear evidence for genetic effects on assortative mating [133].

Ancestry is a particularly important determinant of assortative mating in modern admixed populations [134, 135]. Studies have shown that individuals in admixed Latin American populations tend to mate with partners that have similar ancestry profiles. For example, partners from both Mexican and Puerto Rican populations have significantly higher ancestry similarities than expected by chance [124, 136]. In addition, a number of traits that have been independently linked to assortative mating show ancestry-specific differences in their expression [137]. Accordingly, ancestry-based mate choice has recently been related to a limited number of physical (facial development) and immune-related (MHC loci) traits [124].

The studies that have uncovered the role of genetic ancestry in assortative mating among Latinos have relied on estimates of global ancestry fractions between mate pairs [124, 136]. Given the recent accumulation of numerous whole genome sequences from admixed Latin American populations – along with genome sequences from global reference populations [70] – it is now possible to characterize local genetic ancestry for individuals from admixed American populations [56, 88, 89]. In other words, the ancestral origins for specific chromosomal regions (haplotypes) can be assigned with high

confidence for admixed individuals [51]. For the first time here, we sought to evaluate the impact of local ancestry on assortative mating in admixed Latin American populations. Since the genetic variants that influence numerous phenotypes have been mapped to specific genomic regions, we reasoned that a focus on local ancestry could help to reveal the specific phenotypic drivers of ancestry-based assortative mating.

Our approach to this question entailed an integrated analysis of local genetic ancestry and the genetic architecture of a variety of human traits thought to be related to assortative mating. Assortative mating results in an excess of homozygosity, whereas disassortative mating yields excess heterozygosity. It follows that assortative (or disassortative) mating based on local ancestry would yield an excess (or deficit) of ancestry homozygosity at specific genetic loci. In other words, for a given population, a locus implicated in ancestry-based assortative mating would be more likely to have the same ancestry at both pairs of haploid chromosomes within individuals than expected by chance. We developed a test statistic – the assortative mating index (AMI) – that evaluates this prediction for individual gene loci, and we applied it to sets of genes that function together to encode polygenic phenotypes. We find evidence of substantial local ancestry-based assortative mating, and far less disassortative mating, for four admixed Latin American populations across a variety of anthropometric, neurological and immune-related phenotypes. Our approach also allowed us to assess the specific ancestry components that drive patterns of assortative and disassortative mating in these populations.

5.3 Results

5.3.1 Global and local genetic ancestry in Latin America

We compared whole genome sequences from four admixed Latin American populations, characterized as part of the 1000 Genomes Project (1KGP) [70] to genome sequences and whole genome genotypes from a panel of 34 global reference populations from Africa, Europe and the Americas (Table 4 and Figure 30). The program ADMIXTURE [73] was used to infer the continental genetic ancestry fractions – African, European and Native American – for individuals from the four Latin American populations (Figure 31). Distributions of individuals' continental ancestry fractions illustrate the distinct ancestry profiles of the four populations (Figure 24). Puerto Rico and Colombia show the highest European ancestry fractions along with the highest levels of three-way admixture. These two populations also have the highest African ancestry fractions, although all four populations have relatively small fractions of African ancestry. Peru and Mexico show more exclusively Native American and European admixture, with Peru having by far the largest Native American ancestry fraction.

Table 4. Human populations analyzed in this study.

Populations are organized into continental groups, for both ancestral and admixed Latin American populations, and the number of individuals in each population and group is shown.

	Dataset[1]	Geographical Source	Short	*n*		Dataset[1]	Geographical Source	*n*
Africa (*n*=547)	1KGP	Esan in Nigeria	ESN	99	Native American (*n*=280)	HGDP	Pima in Mexico	14
	1KGP	Gambian in Western Division, The Gambia	GWD	113		HGDP	Maya in Mexico	21
	1KGP	Luhya in Webuye, Kenya	LWK	99		Reich et al.	Tepehuano in Mexico	25
	1KGP	Mende in Sierra Leone	MSL	85		Reich et al.	Mixtec in Mexico	5
	1KGP	Yoruba in Ibadan, Nigeria	YRI	108		Reich et al.	Mixe in Mexico	17
	HGDP	Mandenka		22		Reich et al.	Zapotec in Mexico	43
	HGDP	Yoruba		21		Reich et al.	Kaqchikel in Guatemala	13
Europe (*n*=471)	1KGP	Finnish in Finland	FIN	99		Reich et al.	Kogi in Colombia	4
	1KGP	British in England & Scotland	GBR	90		Reich et al.	Waunana in Colombia	3
	1KGP	Iberian populations in Spain	IBS	107		Reich et al.	Embera in Colombia	5
	1KGP	Toscani in Italy	TSI	107		Reich et al.	Guahibo in Colombia	6
	HGDP	Russian		25		Reich et al.	Piapoco in Colombia	7
	HGDP	Orcadian		15		Reich et al.	Inga in Colombia	9
	HGDP	French		28		Reich et al.	Wayuu in Colombia	11
Admixed (*n*=347)	1KGP	Colombian in Medellin, Colombia	CLM	94		HGDP	Karitiana in Brazil	14
	1KGP	Peruvian in Lima, Peru	PEL	85		HGDP	Suruí in Brazil	8
	1KGP	Mexican Ancestry in LA, California	MXL	64		Reich et al.	Ticuna in Brazil	6
	1KGP	Puerto Rican in Puerto Rico	PUR	104		Reich et al.	Quechua in Peru	40
						Reich et al.	Aymara in Bolivia	23
						Reich et al.	Guarani in Paraguay	6

[1] 1KGP = 1000 Genomes Project; HGDP = Human Genome Diversity Panel; Reich et al. [100]

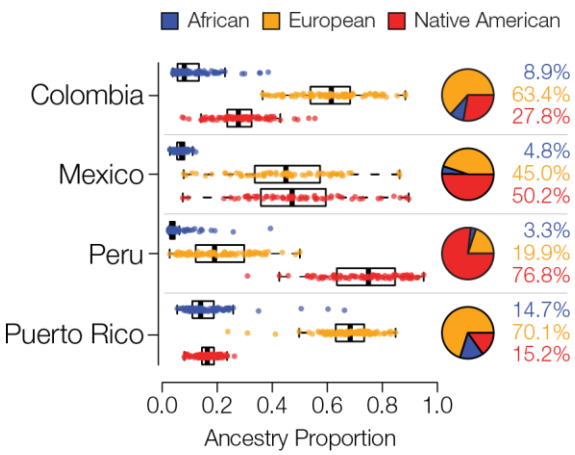

Figure 24. Genetic ancestry proportions for the admixed Latin American populations analyzed here.

For each population, distributions and average values are shown for African (blue), European (orange) and Native American (red) ancestry.

The program RFMix [51] was used to infer local African, European and Native American genetic ancestry for individuals from the four admixed Latin American populations analyzed here. RFMix uses global reference populations to perform chromosome painting, whereby the ancestral origins of specific haplotypes are characterized across the entire genome for admixed individuals. Only haplotypes with high confidence ancestry assignments (≥99%) were taken for subsequent analysis. Examples of local ancestry assignment chromosome paintings for representative admixed individuals from each population are shown in Figure 32. The overall continental ancestry fractions for admixed genomes calculated by global and local ancestry analysis are highly correlated, and in fact virtually identical, across all individuals analyzed here, in support of the reliability of these approaches to ancestry assignment (Figure 33).

5.3.2 Assortative mating and local ancestry in Latin America

We analyzed genome-wide patterns of local ancestry assignment to assess the evidence for assortative mating based on local ancestry in Latin America (Figure 25A). For each individual, the ancestry assignments for pairs of haplotypes at any given gene were evaluated for homozygosity (*i.e.*, the same ancestry on both haplotypes) or heterozygosity (*i.e.*, different ancestry on both haplotypes) (Figure 25B). For each gene, across all four populations, the observed values of ancestry homozygosity and heterozygosity were compared to expected values in order to compute gene- and population-specific assortative mating index (AMI) values. AMI is computed as a log odds ratio as described in the Materials and Methods. The expected values of local ancestry homozygosity and heterozygosity used for the AMI calculations are based on a Hardy-Weinberg (HW) triallelic model with the three allele frequencies computed as the locus-specific ancestry fractions. High positive AMI values result from an excess of observed local ancestry homozygosity and are thereby taken to indicate assortative mating based on shared local genetic ancestry. Conversely, low negative AMI values indicate excess local ancestry heterozygosity and disassortative mating. We performed a series of controls to validate the performance of the AMI test statistic and the justification of the HW model for locus-specific ancestry. These controls are described in detail in *section 5.6.1 Controls for evaluating the assortative mating index (AMI)*, and results of the controls can be seen in Figures 34-36.

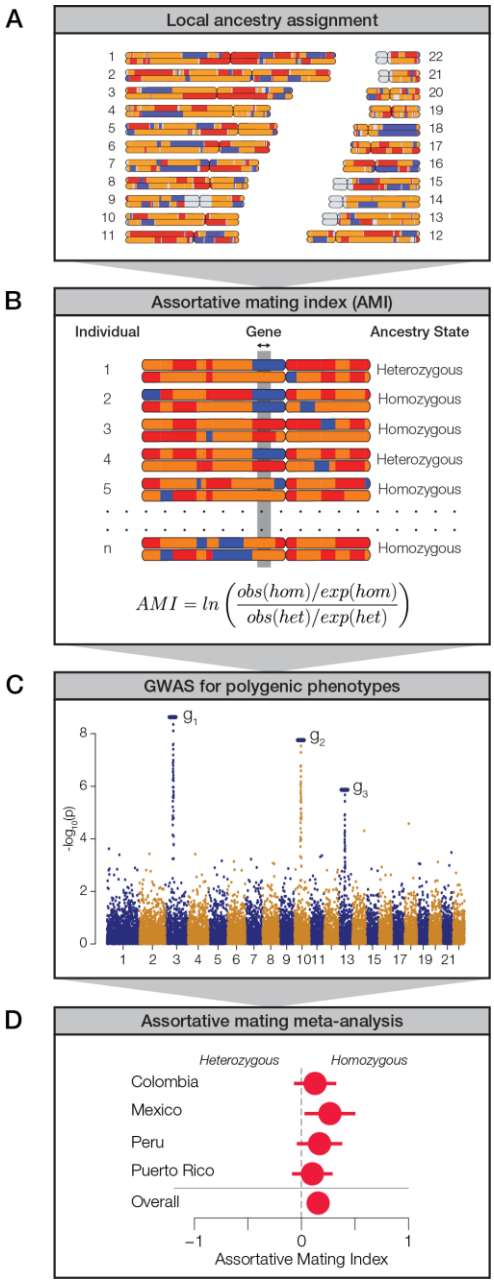

Figure 25. Approach used to measure assortative mating on local ancestry.

(A) Local ancestry is assigned for specific haplotypes across the genome: African (blue), European (orange), and Native American (red). (B) Within individual genomes, genes are characterized as homozygous or heterozygous for local ancestry. For any given population, at each gene locus, the assortative mating index (AMI) is computed from the observed and expected counts of homozygous and heterozygous gene pairs. (C) Data from genome-wide association studies (GWAS) are used to evaluate polygenic phenotypes. (D) Meta-analysis of AMI values is used to evaluate the significance of ancestry-based assortative mating for polygenic phenotypes.

While we were interested in exploring the relationship between local genetic ancestry and assortative mating, we recognized that mate choice is based on phenotypes rather than genotypes *per se*. Since phenotypes are typically encoded by multiple genes, expressed in the context of their environment, we used data from genome-wide association studies (GWAS) to identify sets of genes that function together to encode polygenic phenotypes (Figure 25C). We combined data from several GWAS database sources in order to curate a collection of 105 gene sets that have been linked to the polygenic genetic architecture of a variety of human traits. These gene sets range in size from 2 to 212 genes and include a total of 923 unique genes (Figure 37). We focused on phenotypes that are known or expected to influence mate choice and thereby impact assortative mating patterns. These phenotypes fall into three broad categories: anthropometric traits (*e.g.*, body shape, stature, and pigmentation), neurological traits (*e.g.*, cognition, personality, and addiction), and immune response (HLA genes). Finally, we used a meta-analysis of the AMI values for the sets of genes that underlie each polygenic phenotype in order to evaluate the impact of local ancestry on assortative mating (Figure 25D).

We compared the distributions of observed AMI values versus those expected under random mating to assess the overall evidence for local ancestry-based assortative mating in Latin America. Expected AMI values were computed via permutation analysis by randomly combining pairs of haplotypes into 10,000 diploid individuals for each population to approximate random mating. The distribution of the expected AMI values under random mating is narrow and centered around 0, whereas the observed AMI values have a far broader distribution and tend to be positive (expected AMI μ=-0.01, σ=0.03, observed AMI μ=0.11, σ=0.14; Figure 26A). When all four admixed Latin American

populations are considered together, the mean observed AMI value is significantly greater than the expected mean AMI under random mating (t=18.14, P=8.12e-56). The same trend can be seen when all four populations are considered separately (Figure 38). Mean observed AMI values vary substantially across populations, with Mexico showing the highest levels of local ancestry-based assortative mating and Puerto Rico showing the lowest (Figure 26B). There is also substantial variation seen for the extent of assortative mating among the three broad functional categories of phenotypes (Figure 26C). Local ancestry-based assortative mating is particularly variable for HLA genes, with high levels of assortative mating seen for Mexico and evidence for disassortative mating seen for Colombia and Puerto Rico. Anthropometric traits tend to show higher levels of local ancestry-based assortative mating across all four populations compared to neurological traits.

5.3.3 Local ancestry-based assortative mating for polygenic phenotypes

When considered together, observed AMI levels are enriched for positive values compared to the expected values based on randomly paired haplotypes, indicative of an overall trend of assortative mating based on local ancestry in admixed Latin American populations (Figure 26A and Figure 39). We evaluated polygenic phenotypes individually to look for the strongest examples of traits linked to local ancestry-based assortative mating and to evaluate traits that show either similar or variable assortative mating trends across populations. We computed AMI values for 105 polygenic phenotypes across the four populations; the expected and observed AMI values for all traits are shown in Figure 39. As can be seen for the overall patterns of assortative mating, individual polygenic phenotypes show more extreme positive (for most cases) and negative (in a few cases) AMI

values in the four admixed Latin American populations than can be expected for randomly mating populations.

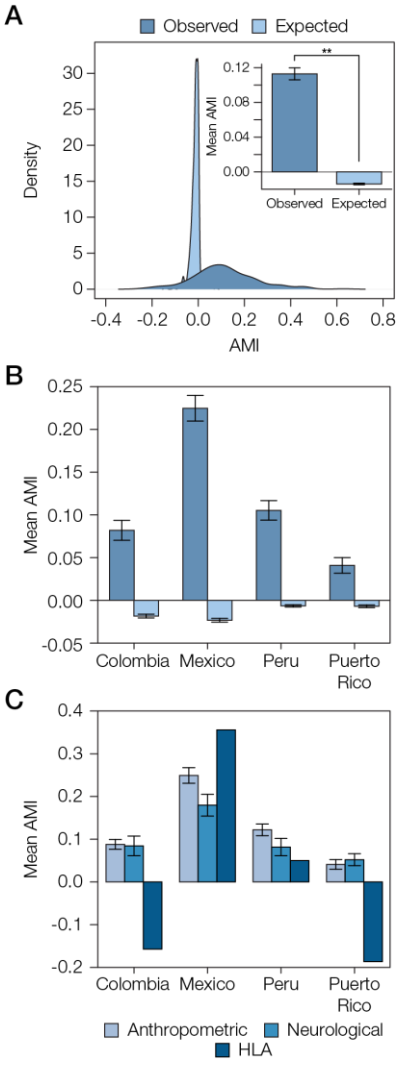

Figure 26. Overview of ancestry-based assortative mating in the four admixed Latin American populations analyzed here.

(A) Distributions of observed and expected AMI values for all four populations. Inset: Mean observed and expected AMI values (±se) for all four populations. Significance between mean observed and expected AMI values (P=8.12e-56) is indicated by two asterisks. (B) Observed and expected average AMI values (±se) across all polygenic phenotype gene sets are shown for each population. (C) Average AMI values (±se) for each population are shown for the three main phenotype functional categories characterized here: anthropometric, neurological, and human leukocyte antigen (HLA) genes.

However, as discussed previously, it is known that admixed Latin American populations mate assortatively based on genetic ancestry [124, 136]. This can be expected to lead to an overall excess of ancestry homozygosity genome-wide compared to expectations based on truly random mating, and we wanted to control for this as well when analyzing the assortative mating signals for specific traits. To do so, we performed an additional permutation analysis by choosing random sets of genes, of the same size as the observed trait-specific gene sets, and then computing AMI values and their significance levels for the randomly permuted gene sets. This allowed us to ask whether the polygenic phenotypes that have statistically significant AMI values show more extreme deviations than can be expected based on genome-wide signals of ancestry-based assortative mating. The results of this additional permutation control are described below in the context of the specific polygenic phenotypes that were found to have significant AMI values.

There are 11 polygenic phenotypes that yield statistically significant AMI values with the HW-based test, after correction for multiple tests, indicative of local ancestry-based assortative mating for specific traits (q<0.05; Figure 27A). A number of other traits shown are marginally significant after correction for multiple testing. We compared the observed AMI values for the gene sets studied here to population-specific null distributions of expected AMI values generated using randomly permuted gene sets as described in the preceding paragraph (Figure 27B). This permutation test accounts for the genome-wide deviations from HW that occur due to ancestry-based assortative mating. The comparisons of the observed versus expected AMI values for this control indicate that the AMI signals that we observe here are trait-specific and cannot be attributed to the elevated levels of ancestry similarity seen for couples in admixed Latin American populations.

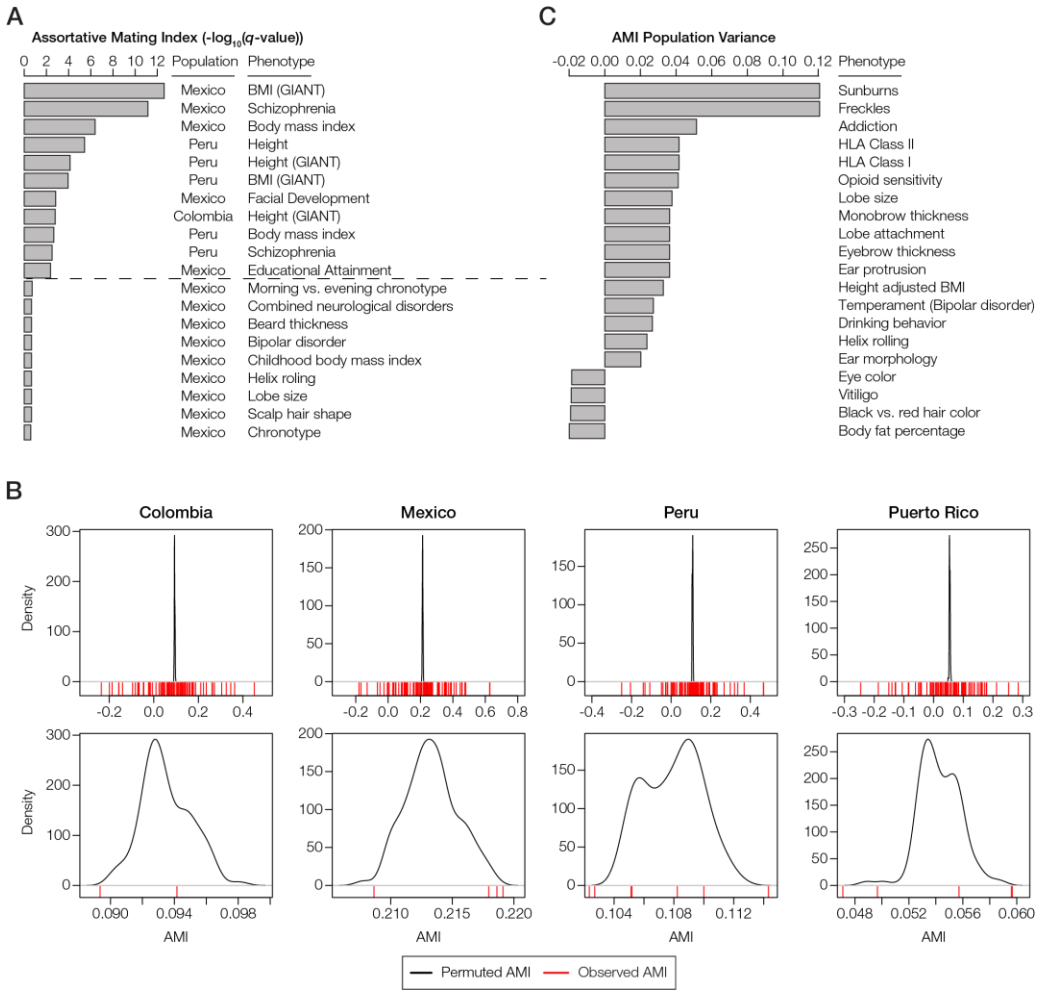

Figure 27. Phenotypes with statistically significant patterns of assortative mating within and among populations.

(A) The top 20 phenotypes with the highest, and most statistically significant, assortative mating values (AMI) seen within any individual population. All AMI values shown are significant at P<0.05, and the dashed line corresponds to a false discovery rate q-value cutoff of 0.05. (B) The observed AMI values for all trait-specific gene sets in each population (red lines) are compared to distributions of expected AMI values (black lines) based on random permutations of 10,000 gene sets. The top panels show the overall distributions of observed and permuted AMI values for each population, with steep peaks around the mean values for expected AMI. The bottom panels show a more narrow range of observed and expected AMI values for each population, which are centered around the population-specific mean expected AMI values. (C) The top 20 phenotypes with the highest or lowest, and most statistically significant, AMI variance levels across populations. Across population variance levels are normalized using the average AMI population variance level for all phenotypes. All AMI variance levels shown are significant at q<0.05. The highest variance (most dissimilar patterns) of the AMI are at the top, while the lowest variance (most similar patterns) of AMI are at the bottom.

The majority of the statistically significant cases of assortative mating are seen in the Mexican and Peruvian populations (10 out of 11), and the anthropometric functional category is most commonly seen among the significant phenotypes (8 out of 11). Height and body mass index are the most commonly observed phenotypes among the significant cases, each appearing in two out of the four populations analyzed here (Colombia and Peru for height and Mexico and Peru for body mass index). The only neurological traits that show significant evidence of assortative mating are schizophrenia (Mexico and Peru) and educational attainment (Mexico). A number of other neurological and anthropometric traits in Mexico are marginally significant. Puerto Rico was the only population that did not show any individual phenotypes with significant evidence of assortative mating, consistent with its low overall AMI values (Figure 26B and Figure 37). A list of these significant traits, including references to the literature where the trait single nucleotide polymorphism (SNP)-associations were originally reported, is provided in Table 7. In addition to evaluating individual phenotypes for statistically significant AMI values, we also looked for polygenic phenotypes that showed the most similar or dissimilar patterns of assortative mating across the four admixed Latin American populations. The top 20 phenotypes with the highest and lowest population variance are shown in Figure 27C (all are statistically significant at $q<0.05$). The polygenic phenotypes with the most variance in population-specific AMI values show more functional diversity compared to the phenotypes with the strongest signals for assortative mating. All three functional categories are represented among the highly population variant phenotypes, and the highly variant phenotypes consist of both assortative and disassortative mating cases (specifically the HLA genes that are described in more detail below). Neurological phenotypes are

enriched among the variant cases, including temperament and several addiction-related phenotypes: opioid sensitivity and drinking behavior. A list of the population (in)variant traits, including references to the literature where the trait SNP-associations were originally reported, is provided in Table 7.

Given the evidence of significant local ancestry-based assortative mating that we observed for a number of traits, we evaluated whether there were particular ancestry components that were most relevant to mate choice. In other words, we asked whether the excess counts of observed ancestry homozygosity or heterozygosity are linked to specific local ancestry assignments: African, European and/or Native American. For significant polygenic phenotype gene sets of interest, we computed the observed versus expected ancestry homozygosity for each ancestry separately across all genes in the set (Figure 28). Height is an anthropometric trait for which Colombia, Mexico, and Peru show significant evidence of assortative mating after correction for multiple tests ($q<0.05$; Figure 28A), and Puerto Rico shows nominally significant assortative mating for this same trait ($P<0.05$). In Colombia, Peru, and Puerto Rico, assortative mating for this polygenic phenotype is driven by an excess of African homozygosity, whereas in Mexico there is a lack of African homozygosity. The neurological disease schizophrenia shows statistically significant assortative mating in Mexico and Peru ($q<0.05$), with marginally significant values in Colombia (Figure 28B). Patterns of assortative mating for this trait in Mexico and Peru are driven mainly by European ancestry, whereas Colombia and Puerto Rico show an excess of African ancestry homozygosity for this same trait.

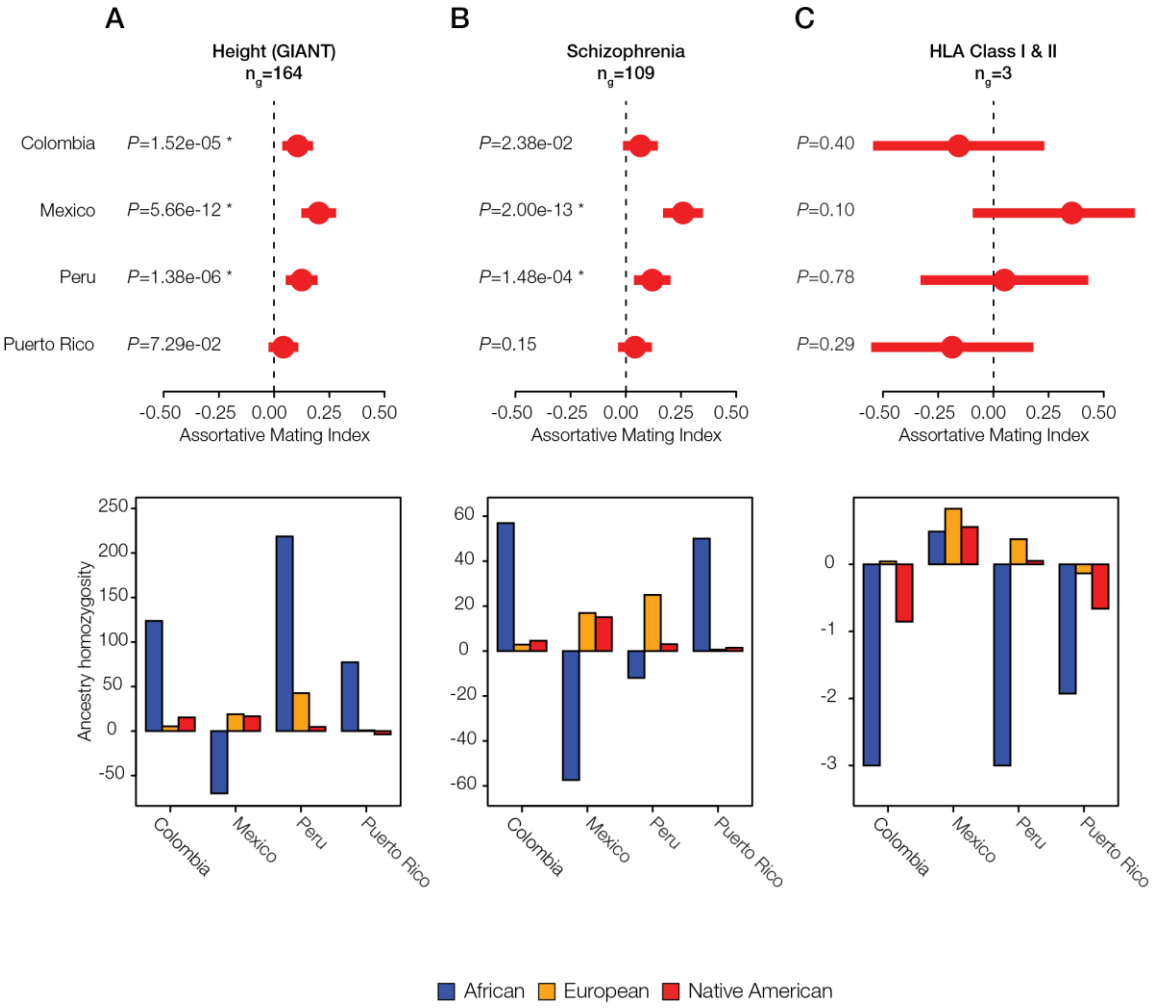

Figure 28. Individual examples of ancestry-based assortative mating and disassortative mating.

Results of meta-analysis of (dis)assortative mating on polygenic phenotypes along with their ancestry drivers are shown for (A) an anthropometric trait: Height, (B) a neurological trait: Schizophrenia, and (C) the immune-related HLA class I and II genes. The meta-analysis plots show pooled AMI odds ratio values along with their 95% CIs and P-values. Stars indicate false discovery rate q-values<0.05. The ancestry driver plots show the extent to which individual ancestry components – African (blue), European (orange), and Native American (red) – have an excess (>0) or a deficit (<0) of homozygosity.

Both Colombia and Puerto Rico show disassortative mating patterns for all HLA loci (both class I and II genes) (Figure 28C). The combined AMI values for the HLA loci are only marginally significant but they are among the lowest AMI values seen for any trait

evaluated here (Figure 39), and they are also highly variable among populations (Figure 27C). HLA loci in Colombia and Puerto Rico show a distinct lack of ancestry homozygosity for almost all ancestry components (Figure 28C). Mexico and Peru, on the other hand, have some evidence for assortative mating for the HLA loci; Mexico has the highest estimates of ancestry homozygosity at HLA loci for any of the four populations, and Peru has an excess of European and Native American ancestry homozygosity and a deficit of African homozygosity for these genes. Similar results for two additional anthropometric phenotypes are shown in Figure 40: body mass index and facial development. These phenotypes show assortative mating in all four populations, with varying components of ancestral homozygosity driving the relationships. When these results are considered together, African ancestry consistently shows the strongest effect on driving assortative and disassortative mating in admixed Latin American populations (Figure 28 and Figure 40).

We further evaluated the extent to which specific ancestry components may drive assortative mating patterns among admixed individuals by evaluating the variance of the three continental ancestry components among individuals within each Latin American population. Assortative mating is known to increase population variance for traits that are involved in mate choice; thus, the ancestry components that drive assortative mating in a given population are expected to show higher overall variance among individual genomes. African ancestry fractions show the highest variation among individuals for all four populations (Figure 29), consistent with the results seen for the five specific cases of assortative mating evaluated in Figure 28 and Figure 40.

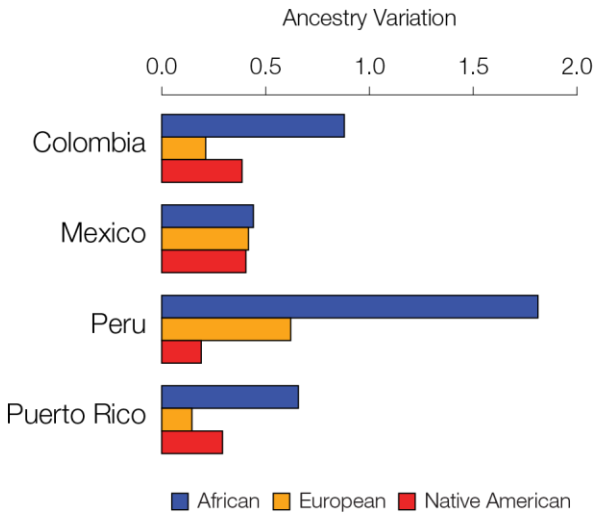

Figure 29. Inter-individual ancestry variance for the four admixed Latin American populations analyzed here.

Variance among individuals for the African (blue), European (orange), and Native American (red) ancestry fractions within each population are shown.

5.4 Discussion

Assortative mating is a nearly universal human behavior, and scientists have long been fascinated by the subject [18, 20]. Studies of assortative mating in humans have most often entailed direct measurements of traits – such as physical stature, education, and ethnicity – followed by correlation of trait values between partners. Decades of such studies have revealed numerous, widely varying traits that are implicated in mate choice and assortative mating. Studies of this kind typically make no assumptions regarding, nor have any knowledge of, the genetic heredity of the traits under consideration. Moreover, the extent to which the expression of these traits varies among human population groups has largely been ignored.

The first attempts to evaluate the genetic contributions to assortative mating entailed twin studies, whereby the similarity of mate choice for dizygotic versus

monozygotic twins were compared [129]. While twin studies did uncover a genetic contribution to variance in human mate choice, they often yielded widely inconsistent results. This was true for both the overall extent of heritability in mate choice, which ranged from 0% to 30% [130, 131], and the relative amounts of genetic versus environmental contributions across the different traits implicated in mate choice [132, 133].

More recent studies of assortative mating, powered by advances in human genomics, have begun to explore the genetic architecture underlying the human traits that form the basis of mate choice in more detail [19, 121]. In addition, recent genomic analyses have underscored the extent to which human genetic ancestry influences assortative mating [124, 135, 136]. However, until this time, these two strands of inquiry have not been brought together. The approach that we developed for this study allowed us to directly assess the connection between local genetic ancestry – *i.e.*, ancestry assignments for specific genome regions or haplotypes – and the human traits that serve as cues for assortative mating.

Previous studies on human mate choice have demonstrated pronounced sex differences in mate preference; for example, females value earning capacity more in potential mates, whereas males value reproductive capacity, as inferred from youth and physical attractiveness [138, 139]. It should be noted that the genome-based approach that we employed here does not allow us to consider sex differences in mate preference since we are essentially observing the effects of assortative mating on the offspring of mate pairs, by comparing ancestry homozygosity levels in the genomes of all individuals, rather than directly observing mate choice in couples.

Our approach relies on the well-established principle that assortative mating results in an excess of genetic homozygosity [134]. However, we do not analyze homozygosity of specific genetic variants *per se*, as is normally done; rather we evaluate excess homozygosity, or the lack thereof, for ancestry-specific haplotypes (Figure 25B). By merging this approach with data on the genetic architecture of polygenic human phenotypes, we were able to uncover specific traits that inform ancestry-based assortative mating. This is because, when individuals exercise mate choice decisions based on ancestry, they must do so using phenotypic cues that are ancestry-associated. In other words, ancestry-based assortative mating is, by definition, predicated upon traits that vary in expression among human population groups (Figure 41). An obvious example of this is skin color [137], and studies have indeed shown skin color to be an important feature of assortative mating [140-143]. It follows that the assortative mating traits that our study uncovered in admixed Latin American populations must be both genetically heritable and variable among African, European and Native American population groups.

The traits we found to influence ancestry-based assortative mating vary among the continental population groups that admixed to form modern Latin American populations (Table 8). For example, the anthropometric traits found in our study – body mass, height and facial development – are both heritable and known to vary among ancestry groups. This implies that the genetic variants that influence these traits should also vary among these populations. Accordingly, it is readily apparent that mate choice decisions based on these physical features could track local genetic ancestry. Interpretation of the neurological traits that show evidence of local ancestry-based assortative mating – schizophrenia and educational attainment – is not quite as straightforward. For schizophrenia, it is far more

likely that we are analyzing genetic loci associated with a spectrum of personality traits that influence assortative mating, as opposed to mate choice based on full-blown schizophrenia, and indeed personality traits are widely known to impact mate choice decisions [119, 122, 125]. In addition, since schizophrenia prevalence does not vary greatly world-wide [144], it is more likely that ancestry-based assortative mating for this trait is tracking an underlying endophenotype rather than the disease itself. While educational attainment outcomes are largely environmentally determined, recent large-scale GWAS studies have uncovered a substantial genetic component to this trait, which is distributed among scores of loci across the genome [145-148]. The population distribution of education-associated variants is currently unknown, but our results suggest the possibility of ancestry-variation for some of them. Indeed, the average allele frequencies for the variants that influence our top four traits of interest – height, body mass index, schizophrenia, and educational attainment – show significant variation among ancestry groups (Figure 41).

Mate choice based on divergent MHC loci, apparently driven by body odor preferences, is the best known example of human disassortative mating [128]. However, studies of this phenomenon have largely relied on ethnically homogenous cohorts. In one case where females were asked to select preferred MHC-mediated odors from males of a different ethnic group, they actually preferred odors of males with more similar MHC alleles [149]. Another study showed differences in MHC-dependent mate choice for human populations with distinct ancestry profiles [127]. Ours is the first study that addresses the role of ancestry in MHC-dependent mate choice in ethnically diverse admixed populations. Unexpectedly, we found very different results for MHC-dependent

mate choice among the four Latin American populations that we studied. In fact, AMI values for the HLA loci are among the most population variable for any trait analyzed here (Figure 27C). Mexico and Peru show evidence of assortative mating at HLA loci, whereas Colombia and Puerto Rico show evidence for disassortative mating (Figure 28C). Interestingly, disassortative mating for HLA loci in Colombia and Puerto Rico is largely driven by African ancestry, and these two populations have substantially higher levels of African ancestry compared to Mexico and Peru. The population- and ancestry-specific dynamics of MHC-dependent mate choice revealed here underscore the complexity of this issue. Given the complexity of the results reported here, particularly as they relate to differences among populations, it should also be noted that anomalous patterns of linkage disequilibrium at MHC loci could confound the analysis at this region.

Assortative mating alone is not expected to change the frequencies of alleles, or ancestry fractions in the case of our study, within a population. Assortative mating does, however, change genotype frequencies, resulting in an excess of homozygous genotypes. Accordingly, ancestry-based assortative mating is expected to yield an excess of homozygosity for local ancestry assignments (*i.e.*, ancestry-specific haplotypes) (Figure 25B). By increasing homozygosity in this way, assortative mating also increases the population genetic variance for the traits that influence mate choice. In other words, assortative mating will lead to more extreme, and less intermediate, phenotypes than expected by chance. This population genetic consequence of assortative mating allowed us to evaluate the extent to which specific continental ancestries drive mate choice decisions in admixed populations, since specific ancestry drivers of assortative mating are expected to have increased variance. We found that the fractions of African ancestry have

the highest variance among individuals for all four populations, consistent with the idea that traits that are associated with African ancestry drive most of the local ancestry-based assortative mating seen in this study (Figure 29).

It is important to reiterate that previous studies have shown evidence for assortative mating on both genetic ancestry and specific traits; accordingly, ancestry-based population stratification could lead to the appearance of trait-based assortative mating [150]. For example, assortative mating occurs among European-Americans along a North-South European ancestry cline, which happens to mirror the cline in height along this same axis [134]. This begs the question as to whether similarities in height among European-American couples is due to ancestry or due to the trait itself. Studies have shown conflicting results regarding this question. On the one hand, genetic ancestry (population stratification) alone was posited to account for observed patterns of assortative mating in the US [151], whereas assortative mating for height was observed within distinct US population groups, independent of their ancestry [133]. Here, we have tried to tease apart the overall effects of ancestry-based assortative mating versus trait-specific mate choice by permuting random sets of genes and re-computing our AMI test statistic for each of the four Latin American populations analyzed here. This procedure allowed us to control for the background levels of local ancestry homozygosity in these populations, which could have been generated by ancestry-based assortative mating alone. We used this control to parameterize the significance levels for the test statistic that we used to discover trait-specific ancestry-based assortative mating (Figure 27B). In other words, we only find a specific trait to be implicated in ancestry-based assortative mating if the levels of ancestry homozygosity for the genes associated with that trait are significantly higher than the

genome-wide background levels. In this sense, we have shown how these specific traits may serve as cues that underlie, to some extent, ancestry influenced mate choice in Latin American populations. A corollary to this conclusion is the fact that ancestry- and trait-based assortative mating cannot be completely disentangled for modern admixed populations. Rather, the variance in the expression of these traits across ancestry groups may in fact be an informative marker for individuals' ancestral origins.

The confluence of African, European and Native American populations that marked the conquest and colonization of the New World yielded modern Latin American populations that are characterized by three-way genetic admixture [56, 65-67, 115]. Nevertheless, mate choice in Latin America is far from random [124, 136]. Indeed, our results underscore the prevalence of ancestry-based assortative mating in modern Latin American societies. The local ancestry approach that we developed provided new insight into this process by allowing us to hone in on the phenotypic cues that underlie ancestry-based assortative mating. Our method also illuminates the specific ancestry components that drive assortative mating for different traits and makes predictions regarding traits that should vary among continental population groups.

5.5 Materials and Methods

5.5.1 Whole genome sequences and genotypes

Whole genome sequence data for the four admixed Latin American populations studied here were taken from the Phase 3 data release of the 1000 Genomes Project (1KGP) [70]. Whole genome sequence data and genotypes for the putative ancestral populations (Africa, Europe, and the Americas) were taken from the 1KGP, the Human Genome

Diversity Project [64] (HGDP), and a previous study on Native American genetic ancestry [100].

Whole genome sequence data and genotypes were merged, sites common to all datasets were kept, and single nucleotide polymorphism (SNP) strand orientation was corrected as needed, using PLINK version 1.9 [71]. The resulting dataset consisted of 1,645 individuals from 38 populations with variants characterized for 239,989 SNPs. The set of merged SNP genotypes was phased, using the program SHAPEIT version 2.r837 [72], with the 1KGP haplotype reference panel. This phased set of SNP genotypes was used for local ancestry analysis. PLINK was used to further prune the phased SNPs for linkage, yielding a pruned dataset containing 58,898 linkage-independent SNPs. This pruned set of SNP genotypes was used for global ancestry analysis.

5.5.2 Global and local ancestry analysis

To infer continental (global) ancestry of the four admixed Latin American populations, ADMIXTURE [73] was run on the pruned SNP genotype dataset (*n*=58,898). ADMIXTURE was run using K=4, yielding African, European, Asian and Native American ancestry fractions of each admixed individual; the final Asian and Native American fractions were summed to determine the Native American fraction of each individual. For local ancestry analysis of the admixed Latin American populations, the program RFMix [51] version 1.5.4 was run in the PopPhased mode with a minimum node size of 5 and the 'usereference-panels-in-EM' option with 2 EM iterations for each individual in the dataset using the phased SNP genotypes (*n*=239,989). Continental African, European, and Native American populations were used as reference populations,

and contiguous regions with the same ancestry assignment, *i.e.*, ancestry-specific haplotypes, were delineated where the RFMix ancestry assignment certainty was at least 99%. The timing of admixture events was analyzed using the program TRACTS [152, 153] with the local ancestry haplotype assignments from RFMix. TRACTS was used to evaluate possible admixture timings across 1,000 bootstrap attempts, with the most likely series of admixture events chosen to represent each population.

Autosomal NCBI RefSeq coding genes were accessed from the UCSC Genome Browser and mapped to the ancestry-specific haplotypes characterized for each admixed Latin American individual. For each diploid genome analyzed here, individual genes can have 0, 1 or 2 ancestry assignments depending on the number of high confidence ancestry-specific haplotypes at that locus. Our assortative mating index (AMI, see below) can only be computed for genes that have 2 ancestry assignments in any given individual, *i.e.*, cases where the ancestry is assigned for both copies of the gene. Thus, for each Latin American population p, the mean ($\overline{x_p}$) and standard deviation (sd_p) of the number of genes with 2 ancestry assignments were calculated and used to compute an ancestry genotype threshold for the inclusion of genes in subsequent analyses. Genes were used in subsequent assortative mating analyses only if they were present above the ancestry genotype threshold of $\overline{x_p} - sd_p$.

5.5.3 Gene sets for polygenic phenotypes

The polygenic genetic architectures of phenotypes that could be affected by assortative mating were characterized using a variety of studies taken from the NHGRI-

EBI GWAS Catalog [107], the Genetic Investigation of ANthropometric Traits (GIANT) consortium[3], and PubMed literature sources.

For each polygenic phenotype, all SNPs previously implicated at genome-wide significance levels of $P \leq 5 \times 10^{-8}$ were collected as the phenotype SNP set. The gene sets for the polygenic phenotypes were collected by directly mapping trait-associated SNPs to genes. SNPs were used to create a gene set only if the SNP fell directly within a gene and thus no intergenic SNPs were used in creating gene sets. Gene sets from the GWAS Catalog were mapped from SNPs using EBI's in-house pipeline. Sets from GIANT were mapped according to specifications of each individual paper. Gene sets from literature searching were mapped using NCBI's dbSNP. For each Latin American population, phenotype gene sets were filtered to only include genes that passed the ancestry genotype threshold, as described previously. Finally, the polygenic phenotype gene sets were filtered based on size, so that all polygenic phenotypes included two or more genes.

Linkage disequilibrium (LD) pruning was performed using the program PLINK to ensure that gene sets consisted of independent genes. LD pruning was done using pairwise r^2 values between genic SNPs for all pairs of genes in any given set. For any pair of genes with SNP $r^2 > 0.1$, only one member of the pair was retained for further analysis. The final data set contains gene sets for 105 polygenic phenotypes, hierarchically organized into three functional categories, including 923 unique genes (haplotypes) (Figure 36).

[3] http://portals.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium

5.5.4 Assortative mating index (AMI)

To assess local ancestry-based assortative mating, we developed the assortative mating index (AMI), a log odds ratio test statistic that computes the relative local ancestry homozygosity compared to heterozygosity for any given gene. Ancestry homozygosity occurs when both genes in a genome have the same local ancestry, whereas ancestry heterozygosity refers to a pair of genes in a genome with different local ancestry assignments. The assortative mating index (AMI) is calculated as:

$$AMI = ln\left(\frac{obs(\text{hom})/\exp(hom)}{\text{obs}(het)/\exp(het)}\right) \quad (9)$$

where $obs(hom)/exp(hom)$ is the ratio of the observed and expected local ancestry homozygous gene pairs and $obs(het)/exp(het)$ is the ratio of the observed and expected local ancestry heterozygous gene pairs.

The observed values of local ancestry homozygous and heterozygous gene pairs are taken from the gene-to-ancestry mapping data for each gene in each population. The expected values of local ancestry homozygous and heterozygous gene pairs are calculated for each gene in a population using a triallelic Hardy-Weinberg (HW) model, in which the gene-specific local ancestry assignment fractions are taken as the three allele frequencies. For the African (a), European (e), and Native American (n) gene-specific local ancestry assignment fractions in a population, the HW expected genotype frequencies are: $(a + e + n)^2$ or $a^2 + 2ae + e^2 + 2an + 2en + n^2$. Accordingly, the expected frequency

of homozygous pairs is $a^2 + e^2 + n^2$ and the expected frequency of heterozygous pairs is $2ae + 2an + 2en$. For each gene, in each population, the expected homozygous and heterozygous frequencies are multiplied by the number of individuals with two ancestry assignments for that gene to yield the expected counts of gene pairs in each class.

For each polygenic phenotype, a meta-analysis of gene-specific AMI values was conducted to evaluate the effect of all of the genes involved in the phenotype on assortative mating, using the metafor [82] package in R. 95% confidence intervals for each gene, meta-gene AMI values, significance *P*-values, and false discovery rate *q*-values, were computed using the Mantel-Haenszel method under a fixed-effects model.

5.5.5 Controls for evaluating the assortative mating index (AMI)

Four different controls were used to evaluate the design and performance of the AMI test statistic: (1) a control for the use of HW as a null model in the AMI test statistic, (2) a permutation analysis to evaluate expected AMI values under random mating, (3) a population genetic simulation to evaluate the power of the AMI test statistic to detect ancestry-based assortative mating and its dependence on the different ancestry combinations of the populations we analyzed, and (4) a permutation of random gene sets to generate null distributions of AMI values expected given the observed genome-wide signals of ancestry-based assortative mating. Each of these control analyses is described in detail in section *5.6 Supplementary Material and Methods*.

5.5.6 Ancestry-specific drivers of assortative mating

For each significant polygenic phenotype of interest, we identified the ancestry component related to mate choice by calculating the ancestry homozygosity ($AH^{anc}_{phenotype}$) for all genes for each ancestry at the given phenotype. The ancestry homozygosity was calculated as

$$AH^{anc}_{phenotype} = \sum_{g \in genes\ in\ the\ phenotype} \left(\frac{obs^{anc}_{g} - exp^{anc}_{g}}{exp^{anc}_{g}} \right) \quad (10)$$

where anc is one of the three ancestries – African, European or Native American, $g \in genes\ in\ the\ phenotype$ are all of the genes involved in the polygenic $phenotype$, obs^{anc}_{g} is the number of observed homozygous genes for gene g coming from anc, and exp^{anc}_{g} is the number of expected homozygous genes for gene g coming from anc (as calculated using a triallelic Hardy-Weinberg model).

5.5.7 Statistical significance testing

Significance testing for the difference between the observed and expected AMI distributions (for both random mating and assortative mating) was completed using the t-test package in R. The metafor package, used for calculating the meta-analysis AMI values, also calculates a *P*-value and a false discovery rate *q*-value to correct for multiple statistical tests, which were used for identifying polygenic phenotypes that are significantly influenced by local ancestry-based assortative mating in each Latin American population. The variance of AMI values across the four populations for each phenotype was calculated

as it is implemented in R and used for identifying phenotypes that had highly similar (minimal variance) or highly dissimilar (maximal variance) local ancestry-based assortative mating patterns. The coefficient of variation was used to measure the inter-individual variance for each of the three continental ancestry components within the four admixed Latin American populations analyzed here.

5.6 Supplementary Material and Methods

5.6.1 Controls for evaluating the assortative mating index (AMI)

Four different controls were used to evaluate the design and performance of the AMI test statistic: (1) a control for the use of HW as a null model in the AMI test statistic, (2) a permutation analysis to evaluate expected AMI values under random mating, (3) a population genetic simulation to evaluate the power of the AMI test statistic to detect ancestry-based assortative mating and its dependence on the different ancestry combinations of the populations we analyzed, and (4) a permutation of random gene sets to generate null distributions of AMI values expected given the observed genome-wide signals of ancestry-based assortative mating. Each of these control analyses is described in the following text.

5.6.1.1 Control 1: Evaluation of Hardy-Weinberg (HW)

HW was intended as a null model against which to test our observed data; nevertheless, it is possible that local ancestry will deviate from HW depending on population history and demography. To control for this possibility, we tested for evidence of (1) the Wahlund effect, which is expected to yield an excess of homozygosity, and (2)

recent admixture, which is expected to yield an excess of heterozygosity. The Wahlund effect, along with the genome-wide distributions of homozygosity and heterozygosity, were measured using the genome-wide distribution of the parameter Φ, where $\Phi = 2pq/\sqrt{p^2 \times q^2}$ with p and q representing ancestry fractions [154]. Genome-wide distributions of heterozygosity were evaluated to look for recent admixture yielding very high heterozygosity. In addition, the admixture timing for the four populations was analyzed using the distributions of the ancestry-specific haplotype lengths with the program TRACTS as described in section *5.5.2 Global and local ancestry analysis*.

5.6.1.2 Control 2: Permutation of random mating

A standard permutation testing framework was adopted for the approximation of random mating in each of the four Latin American populations. Random mating was approximated by randomly combining pairs of individual phased haplotypes from a population to yield permuted diploid genotypes. Haploid chromosomes were permuted randomly within each population using the Fisher-Yates shuffle. After permutation of the chromosomes, per gene AMI values were re-calculated for all genes passing the population-specific ancestry genotyping thresholds. The permutations were completed 20 times, and the population-specific mean AMI values for each gene were taken as the permuted AMI for the gene. This mean permuted AMI per gene was used in AMI meta-analysis for each gene set to determine expected AMI values.

5.6.1.3 Control 3: Population genetic simulation of assortative mating

To validate the performance of the AMI test statistic, we adopted a population genetic model that simulates assortative mating in the four Latin American populations

under Hardy-Weinberg equilibrium, with a fraction of the population mating assortatively. For each gene in a given population, the present-day local ancestry assignment fractions are used as the starting ancestral proportions: African = a, European = e, Native American = n. Using a triallelic Hardy-Weinberg model, taking the ancestral proportions as the allele frequencies, the ancestry genotype frequencies for a given gene at the starting generation are calculated as:

$$P_{aa} = a^2$$

$$P_{ae} = 2ae$$

$$P_{an} = 2an$$

$$P_{ee} = e^2$$

$$P_{en} = 2en$$

$$P_{nn} = n^2$$

where P_{aa} = African-African genotype, P_{ae} = African-European genotype, P_{an} = African-Native American genotype, P_{ee} = European-European genotype, P_{en} = European-Native American genotype and P_{nn} = Native American-Native American genotype. Under the model, the fraction of the population that mates assortatively is denoted as α and the fraction that mates randomly is $1 - \alpha$. Taking the current generation ancestry genotype frequencies, the subsequent generation's ancestry genotype frequencies are calculated using the formulae:

$$P'_{aa} = (1 - \alpha) \times a^2 + \alpha \times (P_{aa} + 0.25 \times P_{ae} + 0.25 \times P_{an})$$

$$P'_{ae} = (1-\alpha) \times 2ae + \alpha \times (0.5 \times P_{ae})$$

$$P'_{an} = (1-\alpha) \times 2an + \alpha \times (0.5 \times P_{an})$$

$$P'_{ee} = (1-\alpha) \times e^2 + \alpha \times (P_{ee} + 0.25 \times P_{ae} + 0.25 \times P_{en})$$

$$P'_{en} = (1-\alpha) \times 2en + \alpha \times (0.5 \times P_{en})$$

$$P'_{nn} = (1-\alpha) \times n^2 + \alpha \times (P_{nn} + 0.25 \times P_{an} + 0.25 \times P_{en})$$

Ancestry genotypes in each population were simulated for 20 generations, with the assumption of a generation time of 25 years, accounting for 500 years of elapsed time during the conquest and colonization of the Americas. The final ancestry genotype frequencies after the 20 generations were used to calculate the simulated ancestry homozygosity and heterozygosity values. For each Latin American simulated population, random gene sets, ranging in size from 2 to 20, were created by subsampling genes in the simulation. A meta-analysis AMI value and *P*-value for each gene set was calculated using the fixed-effects model of the Mantel-Haenszel method.

5.6.1.4 Control 4: Permutation test of ancestry-based assortative mating

Permutation of random gene sets was used to generate null distributions of gene set AMI values expected given the observed genome-wide levels of ancestry-based assortative mating. This permutation controls for the genome-wide levels of ancestry homozygosity based on overall levels of ancestry similarity between couples in admixed Latin American populations. For these permutations, 10,000 sets of genes, of the same sizes as the gene sets curated for the polygenic traits analyzed here, were permutated by randomly selecting

genes without replacement from across the genome, and AMI values for the random gene sets were calculated. The randomly permuted AMI distributions were compared to the observed AMI values for the significant polygenic trait gene sets to evaluate the extent to which observed trait gene set AMI values deviate from AMI values seen for random gene sets.

5.7 Supplementary Results

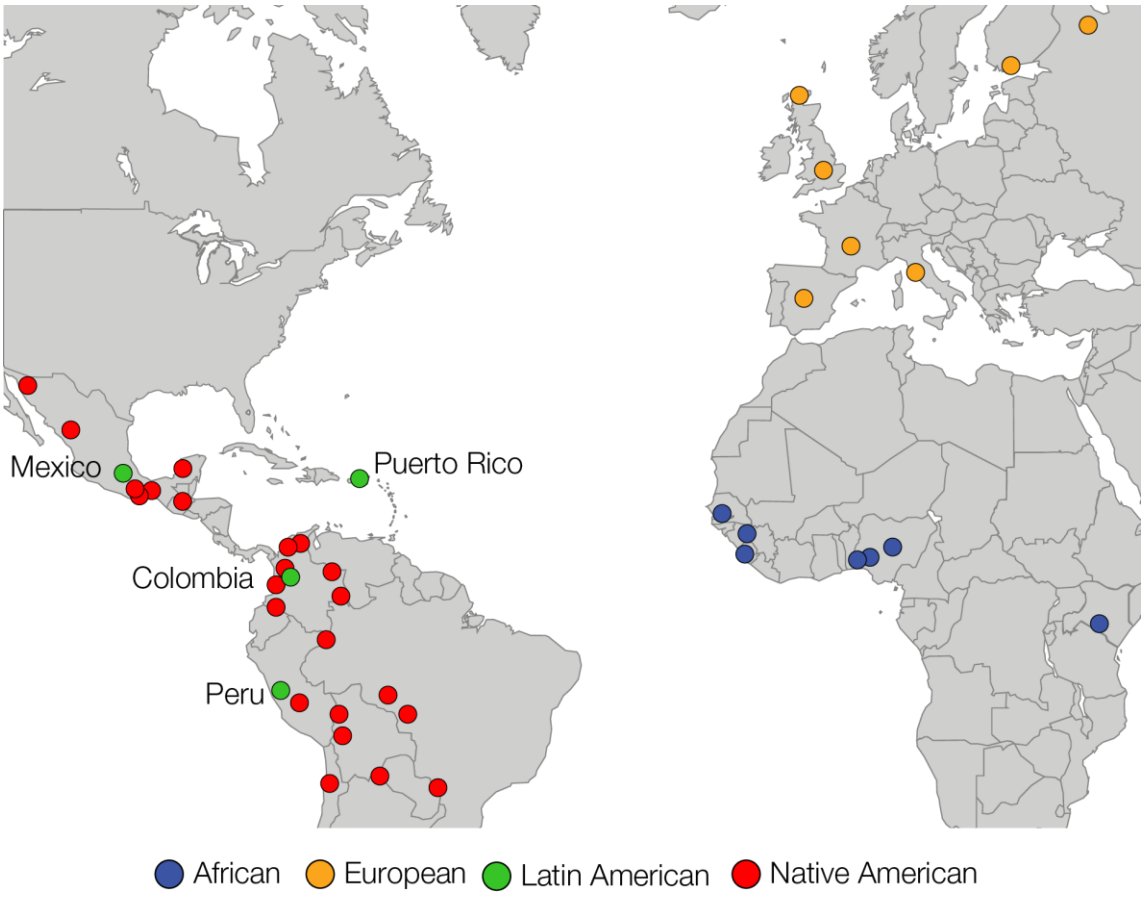

Figure 30. Global locations of the populations analyzed in this study.

Global reference populations – African (blue), European (orange) and Native American (red) – were used to infer the continental ancestry proportions of the four admixed Latin American populations (green) studied here. Map adapted from https://commons.wikimedia.org/wiki/File:World_map_(Miller_cylindrical_projection,_blank).svg

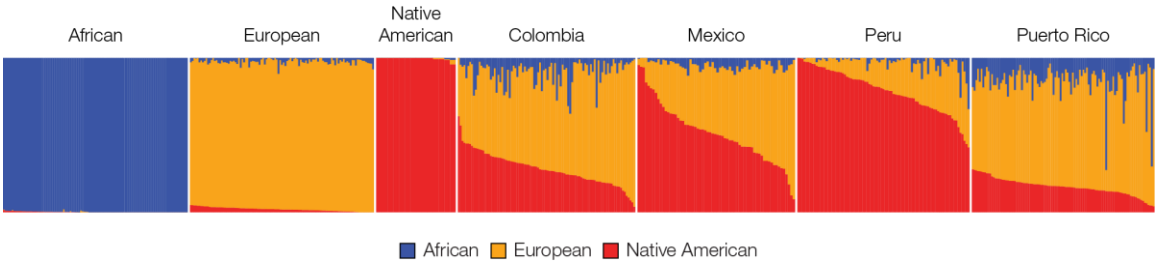

Figure 31. Three-way continental genetic ancestry for the four admixed Latin American populations analyzed in this study.

ADMIXTURE plot showing genome-wide continental ancestry fractions for each individual in each of the four Latin American populations and for each individual in the global reference populations: African (blue), European (orange) and Native American (red).

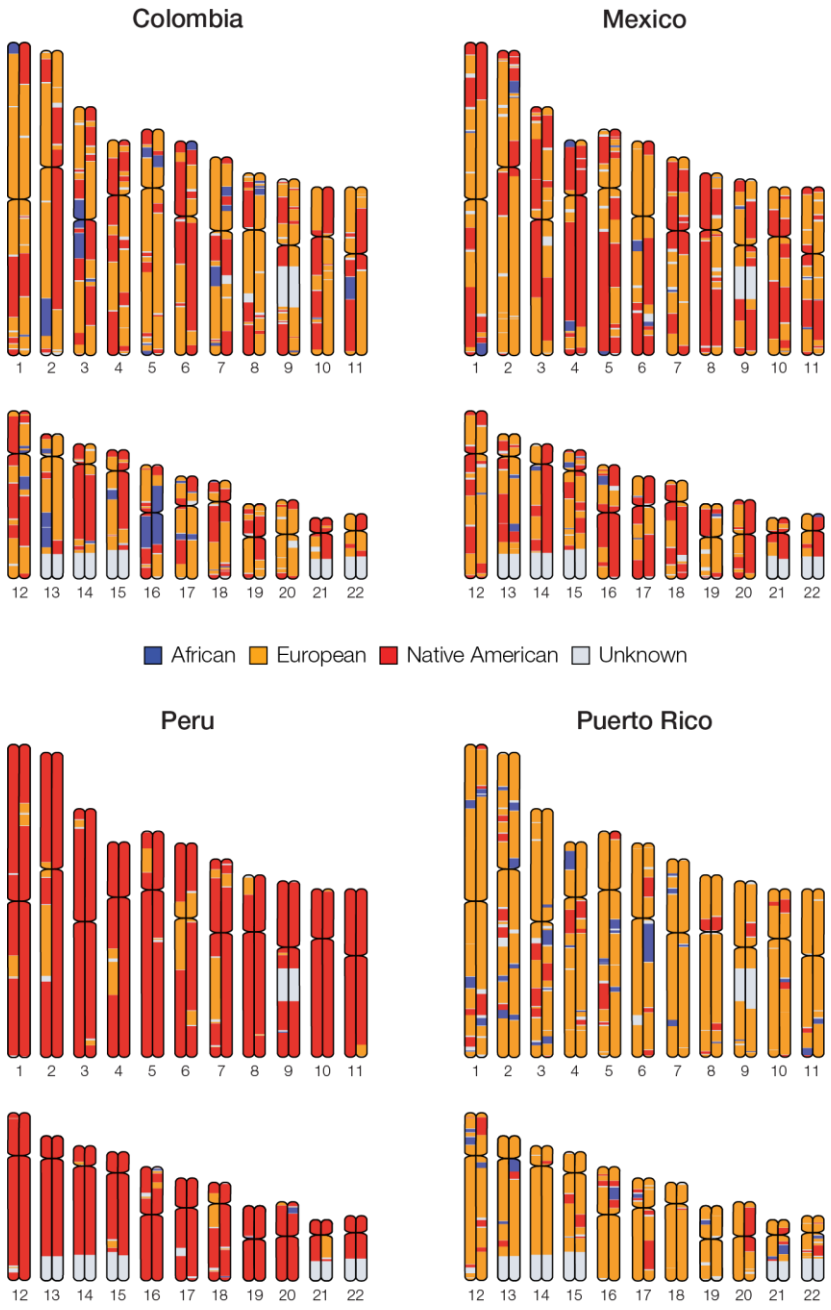

Figure 32. Local ancestry assignment with chromosome painting.

Examples of local ancestry assignment chromosome paintings are shown for Colombia, Mexico, Peru and Puerto Rico. Examples correspond to genomes that have close to average ancestry proportions for each population.

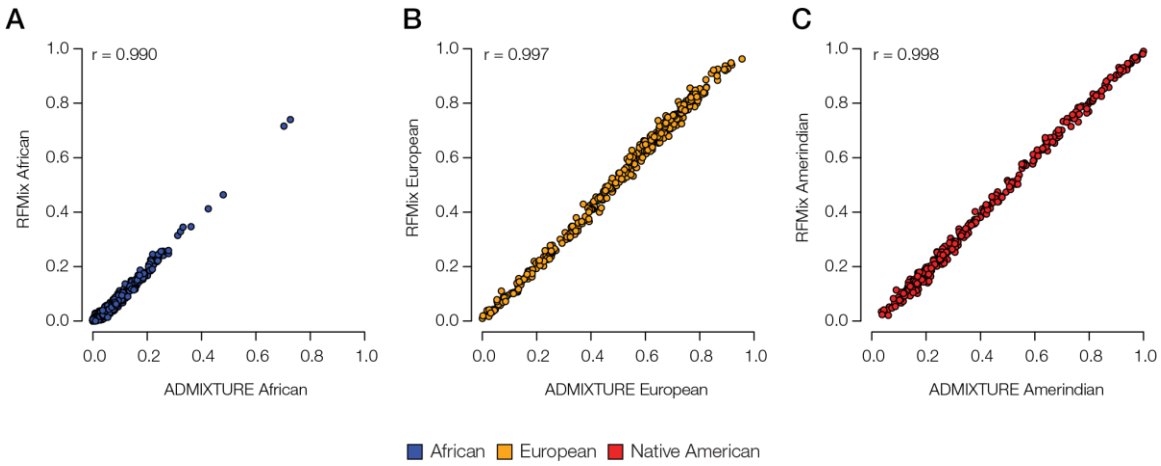

Figure 33. Comparison of ancestry fractions estimated by ADMIXTURE (global ancestry) versus RFMix (local ancestry).

*Genome-wide continental ancestry fractions were inferred for all individuals using the program ADMIXTURE, which estimates global ancestry fractions, and by summing the lengths of local ancestry-specific tracts (haplotypes) inferred from RFMix. Correlations between ADMIXTURE (x-axis) and RFMix (y-axis) ancestry estimates for individuals analyzed here are shown for (**A**) African (blue), (**B**) European (orange), and (**C**) Native American (red) ancestry. Pearson correlation r-values are shown for each ancestry.*

5.7.1 Control 1: Evaluation of Hardy-Weinberg (HW)

We chose HW as an idealized population genetic model against which to test the observed distributions of local ancestry. Nevertheless, genome-wide patterns of local ancestry may deviate from HW depending on population history and demography. On the one hand, population structure may lead to the so-called Wahlund effect, yielding a genome-wide excess of homozygosity. On the other hand, very recent admixture could lead to a genome-wide excess of heterozygosity. We controlled for these two possibilities to validate the use of HW as a null model for the genome-wide distribution of local ancestry. Genome-wide patterns of homozygosity and heterozygosity were measured using the parameter Φ for the four admixed Latin American populations. Φ is expected to

equal 2 under HW, whereas $\Phi<2$ indicates an excess of homozygosity and $\Phi>2$ indicates an excess of heterozygosity [154]. All four populations show genome-wide median values of Φ very close to 2 as well as genome-wide distributions centered around 2, in support of the use of HW as a null model for local ancestry distribution (Figures 34A and 34B).

We also computed the genome-wide median values and distributions of heterozygosity to control for the possibility of recent admixture yielding extremely high values of ancestry heterozygosity genome-wide. Median values of heterozygosity are close to 0.5 for all four populations analyzed, and the genome-wide heterozygosity distributions do not show any evidence for extreme ancestry heterozygosity values caused by recent admixture (Figures 34C and 34D). We further evaluated admixture timing for the four populations via analysis of the distribution of ancestry-specific haplotype lengths with the program TRACTS. The TRACTS analysis does not show any evidence for recent admixture among these populations; the inferred admixture events range from 8 to 15 generations ago, with an average of ~11 generations (Figure 35). These estimates are consistent with previous studies as well as the known history of the region [85, 88, 89]. Taken together, the results of the heterozygosity and admixture timing analyses also support the use of HW as a null model for local ancestry distribution.

5.7.2 Control 2: Permutation of random mating

Results of the chromosome permutation control for random mating are shown as the expected AMI distributions in Figures 26A and 26B (see section *5.3 Results* for further description).

5.7.3 Control 3: Population genetic simulation of assortative mating

In addition to the permutation test, we also performed a simulation analysis using a population genetic model of assortative mating to assess the power of the AMI test statistic (Figure 36). We were particularly interested in exploring the potential effects of different ancestry proportions among the populations analyzed here, and different gene set sizes, on computed AMI values. The population genetic simulation shows that our AMI test statistic is sensitive even when the fraction of the population that mates assortatively is low. We also found that AMI values are not biased in any particular direction based on the overall ancestry fractions observed for each population. For example, according to the AMI power simulation, Colombia should have the highest overall AMI values, followed by Puerto Rico, Mexico, and Peru. This order is completely different from what is seen for the observed AMI values, where Mexico shows the highest mean value, followed by Peru, Colombia, and Puerto Rico (Figure 26B). The population genetic simulation does show that the size of the gene set being analyzed influences the sensitivity of the AMI test statistic. Larger gene sets show greater evidence for assortative mating at the same α parameter values compared with smaller gene sets.

5.7.4 Control 4: Permutation test of ancestry-based assortative mating

Results of the random genet set permutation control for assortative mating are shown as the permuted AMI distributions in Figure 27B (see section *5.3 Results* for further description).

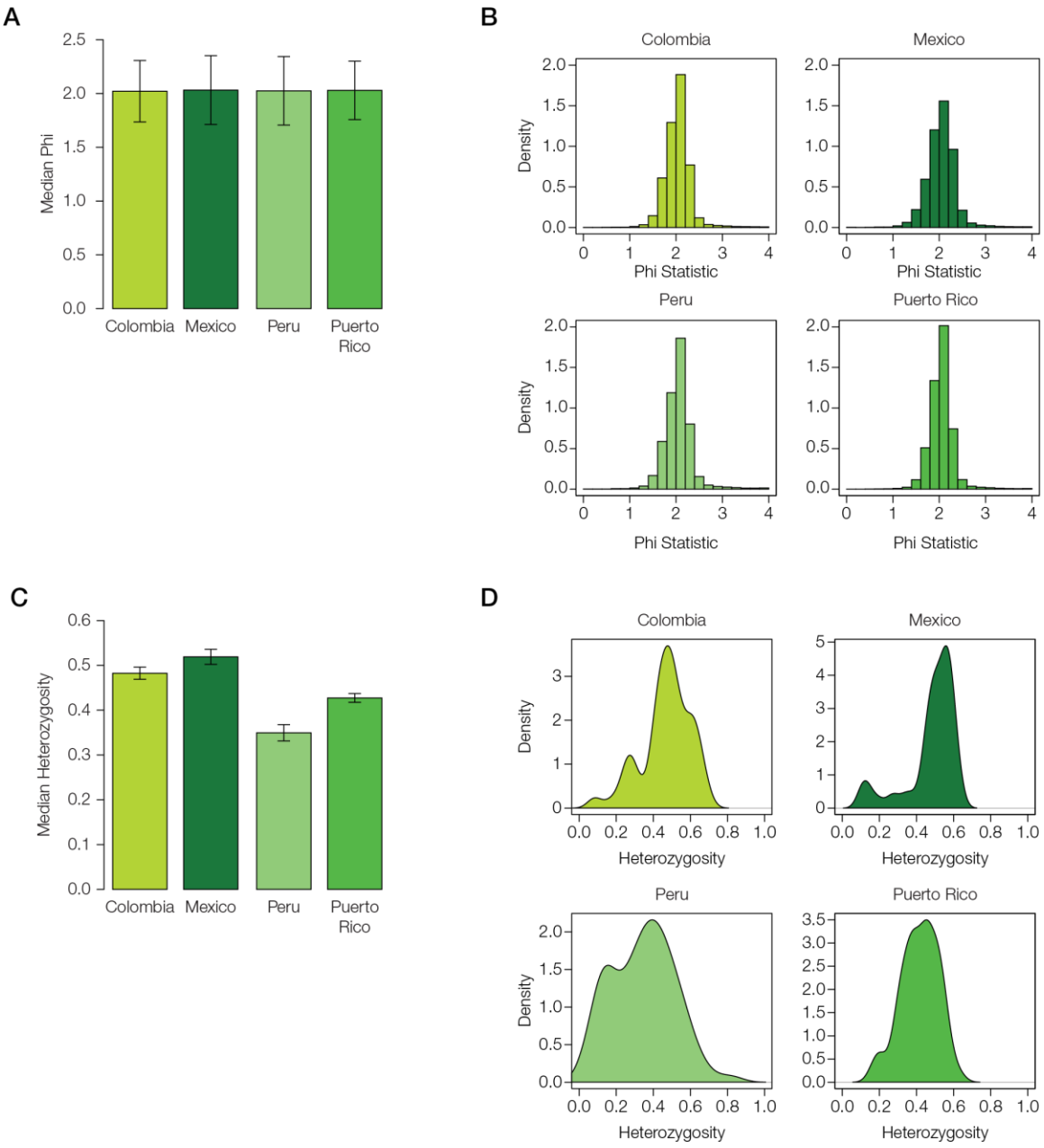

Figure 34. Genome-wide patterns of homozygosity and heterozygosity for the four admixed Latin American populations analyzed in this study.

Genome-wide median values (A) and distributions (B) for the parameter Phi (Φ) were used to test for excess homozygosity and heterozygosity genome-wide. Genome-wide median values (C) and distributions (D) of heterozygosity were used to test for very recent admixture, which are expected to yield ~100% heterozygosity genome-wide.

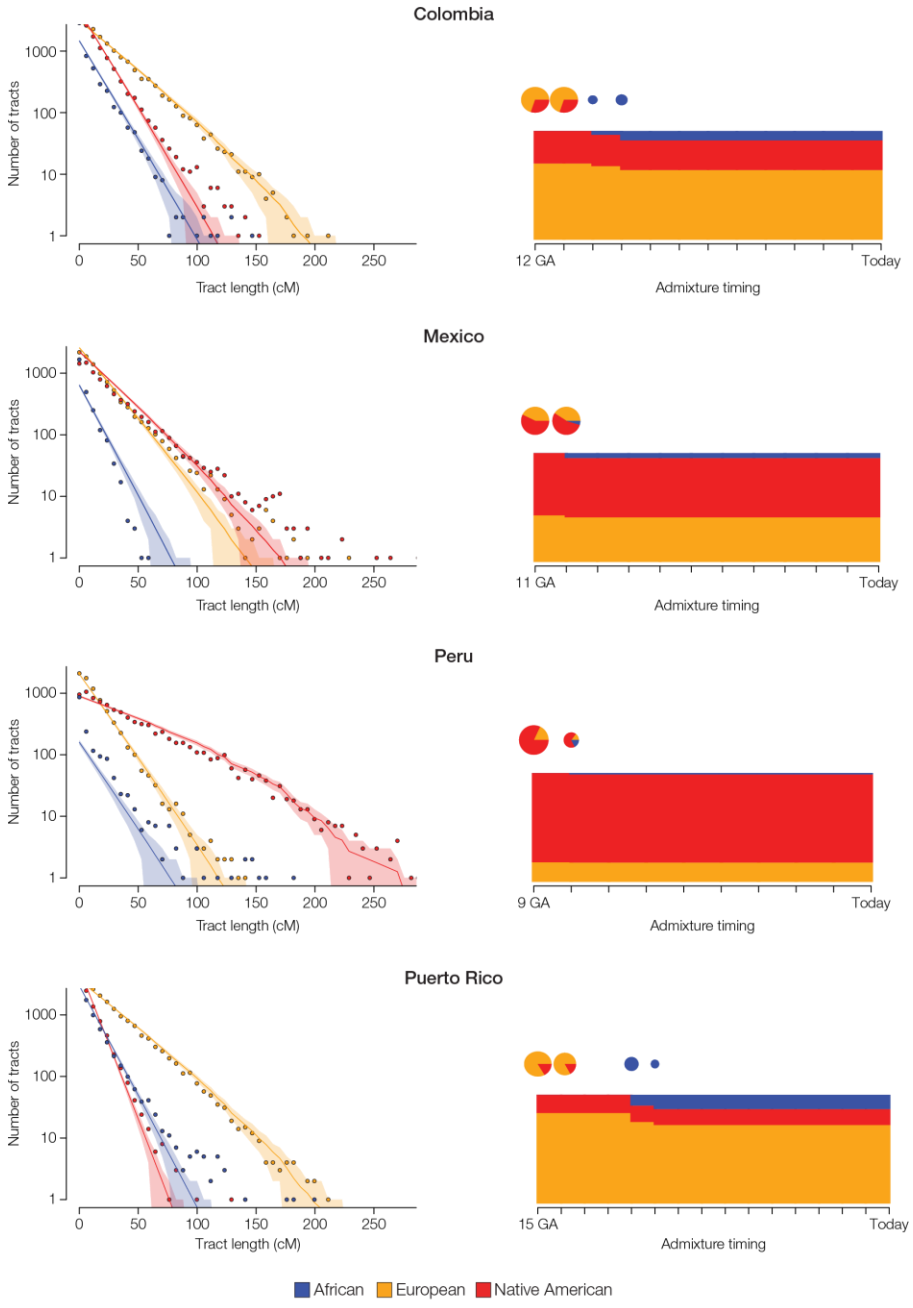

Figure 35. Admixture timing for the four admixed Latin American populations analyzed in this study.

(Left panels) Observed (points) and predicted (solid lines) ancestry tract size distributions, with shaded areas representing 95% confidence intervals. (Right panels) Admixture event timings are shown together with ancestry proportions. Each inferred admixture event is indicated by a circle, which is scaled according to the size of the contribution to the population and also shows the relative ancestry proportions. The y-axes of the charts show the inferred continental ancestry fractions, and the x-axes show admixture timing as the number of generations ago (GA).

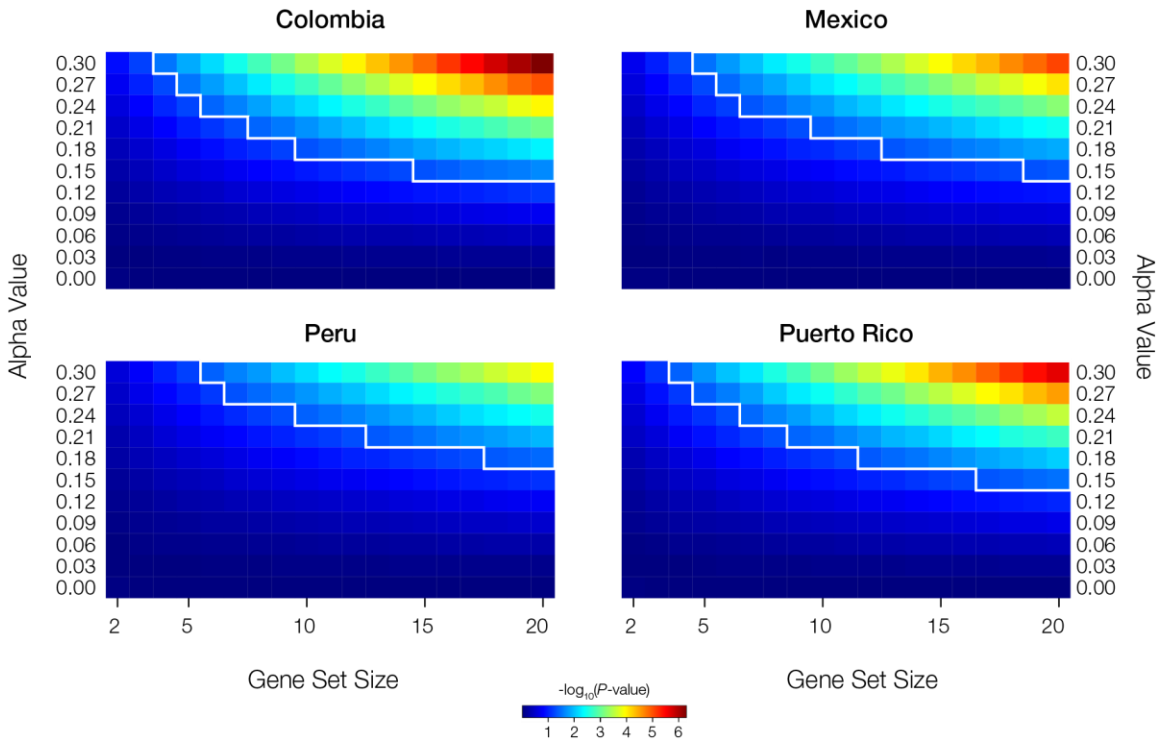

Figure 36. Simulation of the assortative mating index (AMI) test statistic under assortative mating.

Assortative mating was modeled combining Hardy-Weinberg genotype expectations with a single parameter α that represents the fraction of the population that mates assortatively. Assortative mating α-values range from 0 (no assortative mating) to 0.3 (incomplete assortative mating) for polygenic phenotypes encoded by gene sets of n_g=2 to 20 genes. For each population, statistical significance P-values for AMI are plotted for all combinations of α and n_g, and the area of significant (P<0.05) AMI values is indicated.

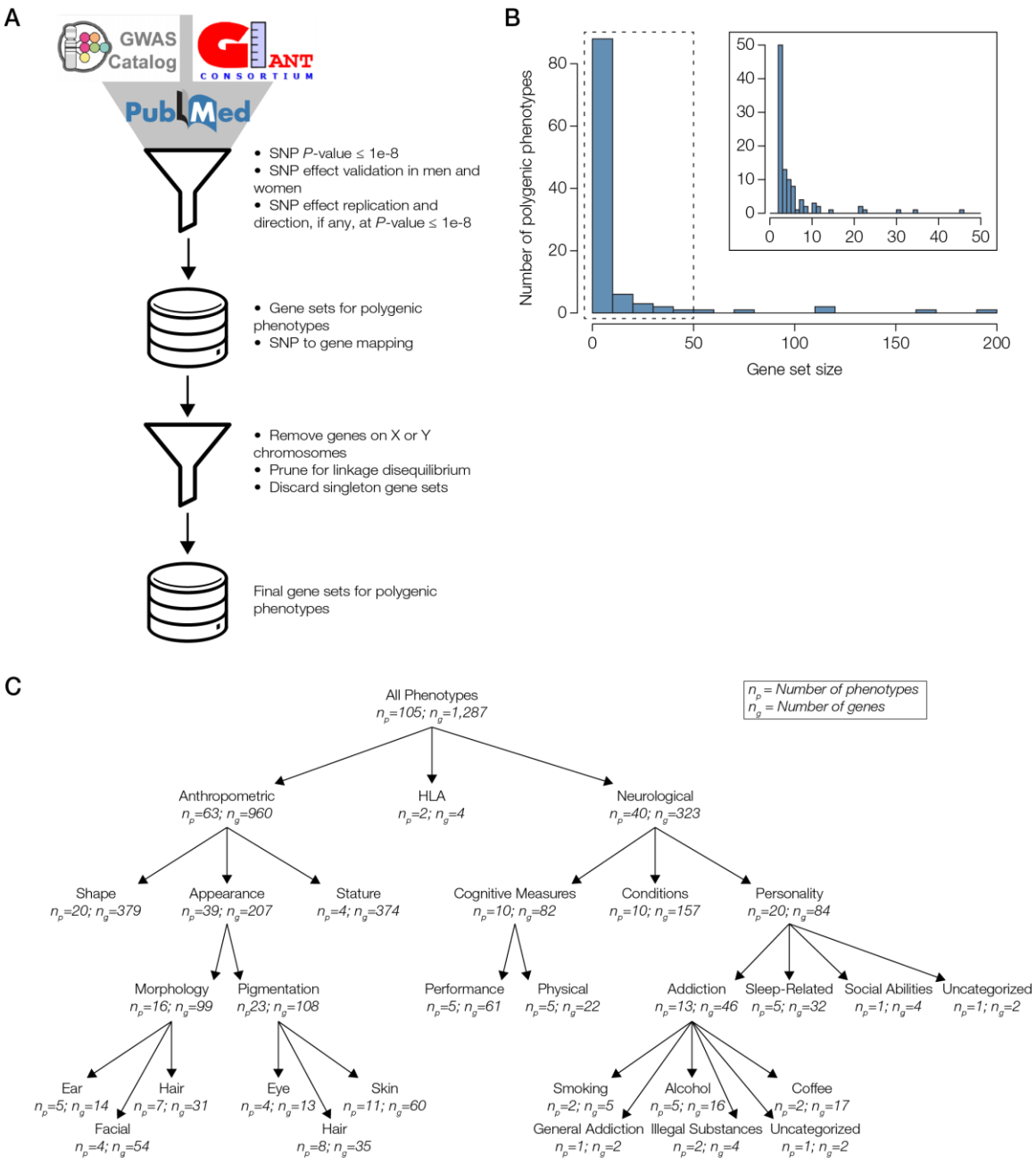

Figure 37. Polygenic phenotypes taken from genome-wide association studies (GWAS).

(A) GWAS SNP-associations curated from (1) the NHGRI-EBI GWAS catalog, (2) the Genetic Investigation of ANthropometric Traits (GIANT) consortium, and (3) scientific literature indexed in PubMed were mapped to genes as described in the Methods in order to evaluate ancestry-based assortative mating on polygenic phenotypes. (B) Distribution of the number of genes per polygenic phenotype (trait). (C) Hierarchical organization scheme developed for the polygenic phenotypes analyzed here. The numbers of phenotypes (n_p) and total genes (n_g) are shown for each node in the scheme tree.

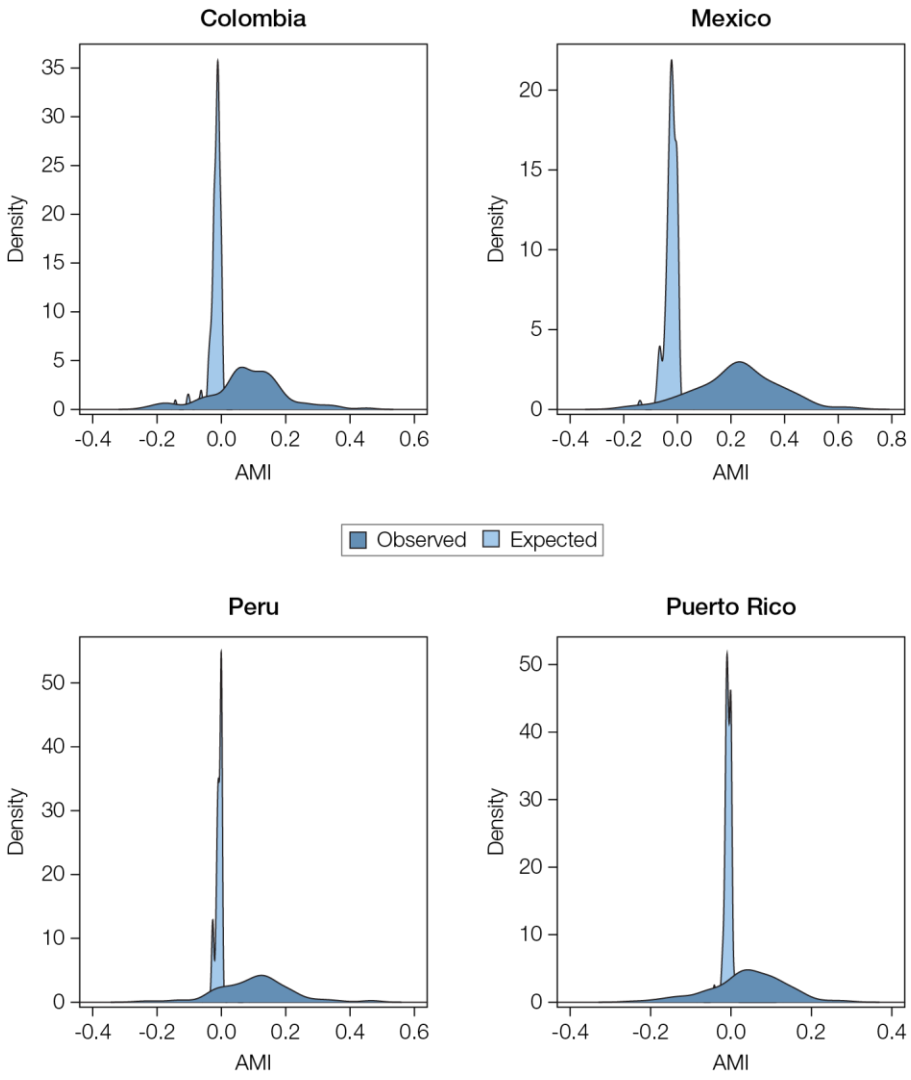

Figure 38. Distributions of observed (dark blue) versus expected (light blue) AMI values for the four admixed Latin American populations analyzed here.

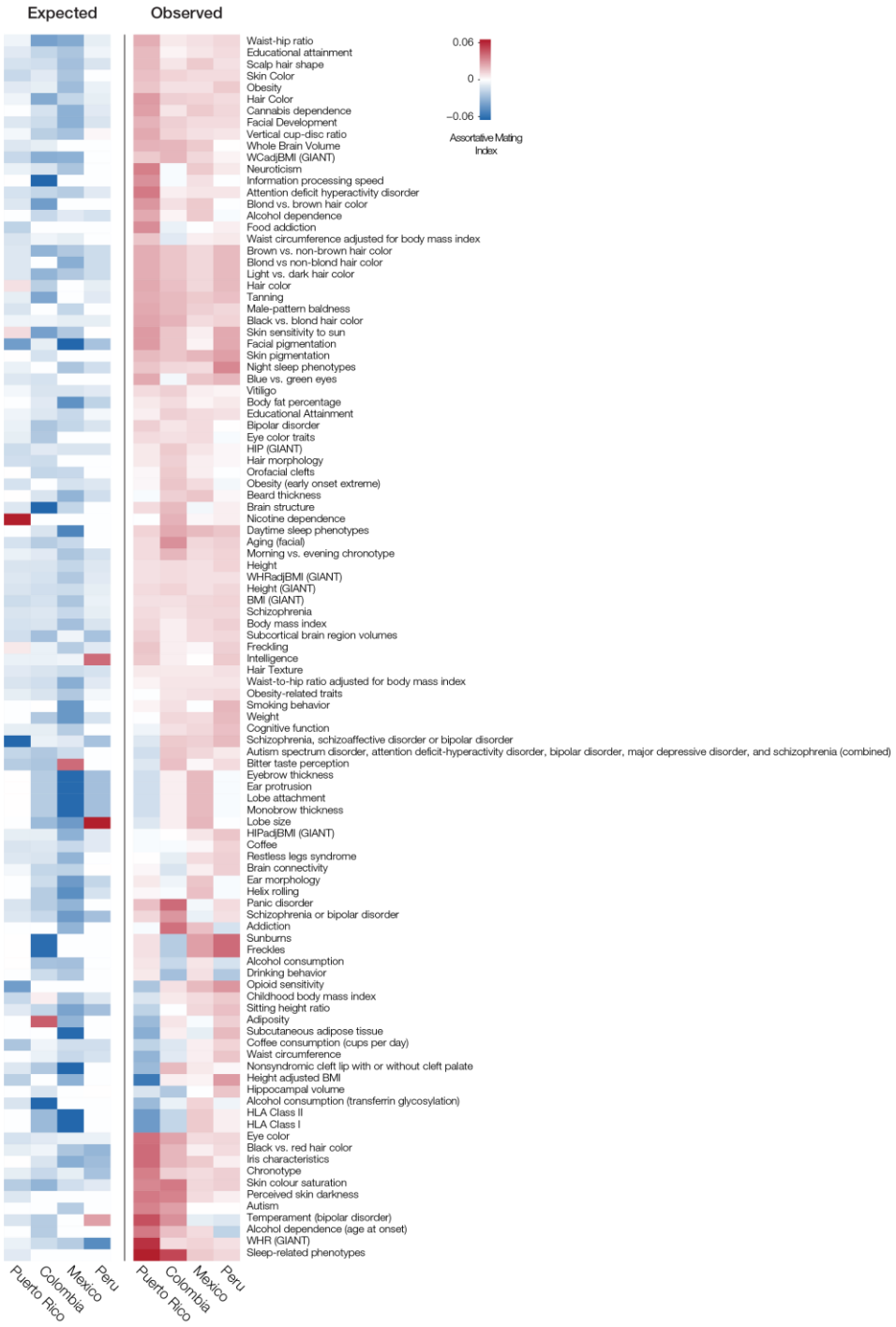

Figure 39. Assortative mating index (AMI) values for all phenotypes across all four populations analyzed here.

AMI values are color coded with red for positive values (corresponding to assortative mating) and blue for negative values (corresponding to disassortative mating). The AMI value matrix is hierarchically clustered on both the y- and x-axes in order to visualize similar and divergent assortative mating trends across populations.

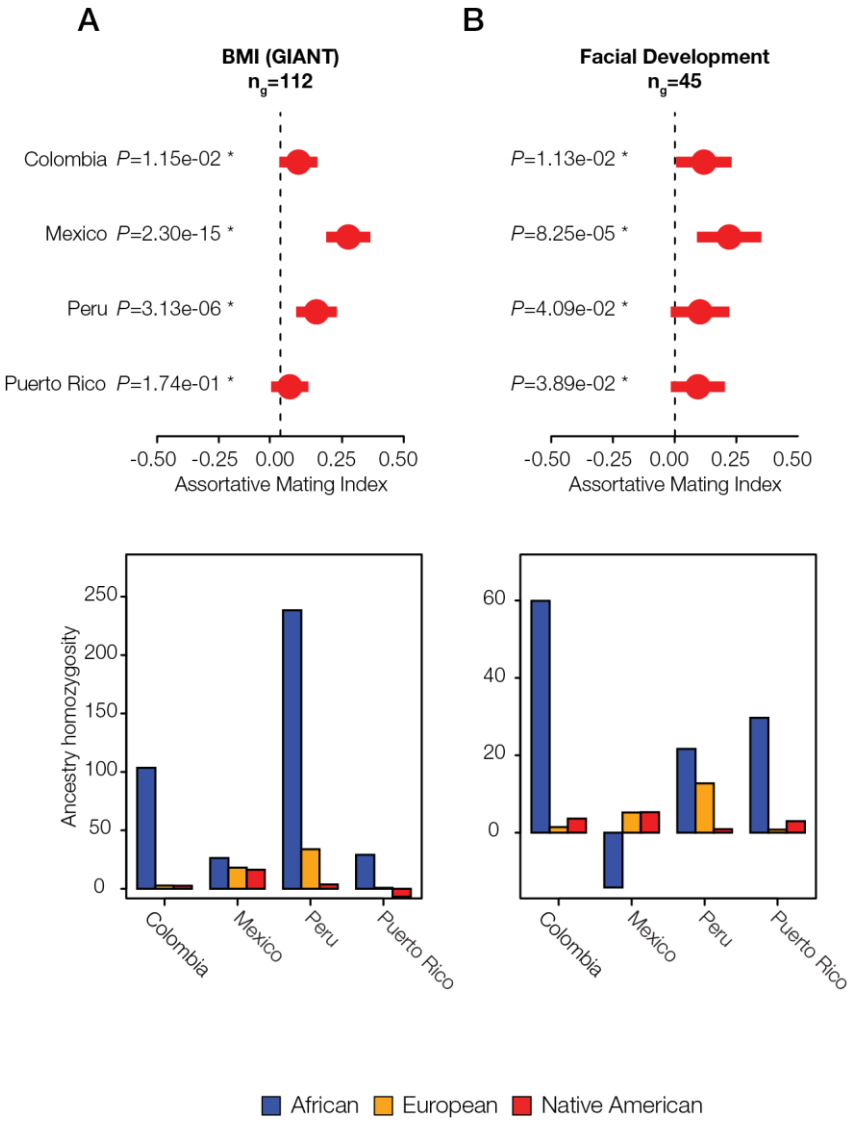

Figure 40. Individual examples of ancestry-based assortative mating.

Results of meta-analysis of (dis)assortative mating of polygenic phenotypes along with their ancestry drivers are shown for (A) body mass index, and (B) facial development. The meta-analysis plots show pooled AMI odds ratio values along with their 95% CIs and P-values. Stars indicate false discovery rate q-values<0.05. The ancestry driver plots show the extent to which individual ancestry components have an excess or deficit of homozygosity.

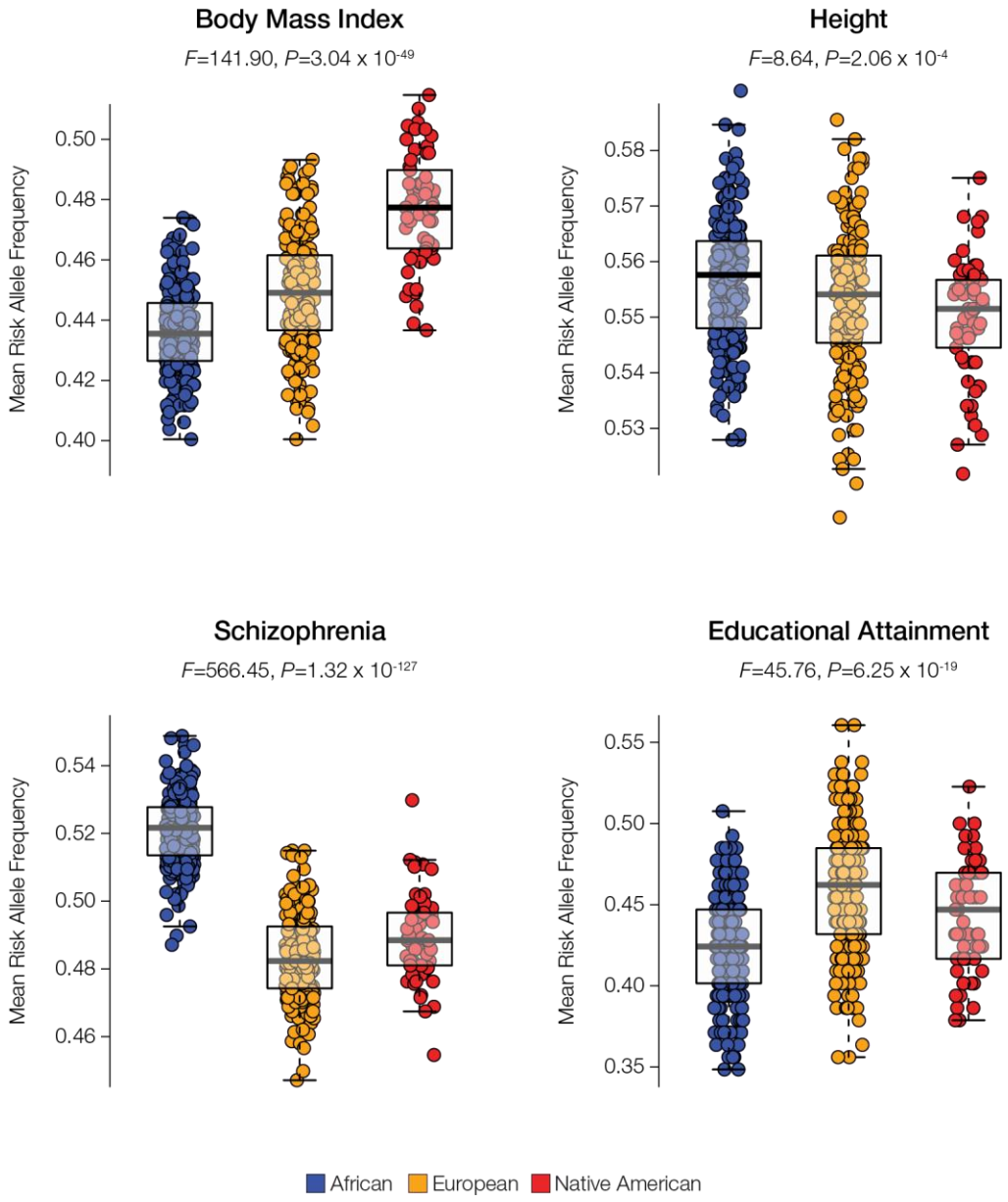

Figure 41. Genetic variation in trait-specific SNP frequencies across continental ancestry groups.

Average risk allele frequency distributions are shown for the top four ancestry-based assortative mating traits (see Figure 27A). Average SNP risk allele frequencies for each trait were computed using the 1000 Genomes Project data for individuals from African (GWD and YRI), European (CEU and IBS), and Native American (MXL and PEL) populations. To approximate Native American ancestry, only Mexican (MXL) and Peruvian (PEL) individuals with >75% Native American ancestry were chosen (see Figure 31). ANOVA was used to evaluate the significance of the differences in the ancestry-specific distributions for each trait (see F- and P-values for each plot).

CHAPTER 6. CONCLUSIONS AND FUTURE PROSPECTS

Admixture is a ubiquitous feature of human evolution, and examples of it can be found throughout human history. No longer is it believed that humans solely evolved by the serial founder model – geographic and reproductive isolation followed by genetic divergence. Instead, it has been shown that the recurrent and joint processes of migration-driven genetic divergence and admixture have been present throughout human evolution. Examples of this cycle are seen as early as anatomically modern humans admixing with Neandertals and Denisovans, up through the Columbian Exchange ~500 years ago, and ongoing today at an ever increasing rate.

The Columbian Exchange brought together once isolated populations in a new environment over a short period of time. The movement of Europeans and Africans into Native American lands in the New World and subsequent admixture created the modern populations of Latin America. Individuals from these populations have three-way admixture and can be considered to possess evolutionarily novel genomes, with pre-adapted ancestral haplotypes coinciding on genomic backgrounds that were never before combined. Latin American populations, while largely sharing the same continental ancestral components, have different levels of admixture among the three ancestries and different patterns of local ancestry as well.

The combinations of genetic ancestry seen in admixed populations play an important role when investigating the genomic determinants of health and disease. Relationships between ancestry and disease-related variants become more complicated to study when looking at admixed populations, since the patterns of local ancestry vary among individuals

and populations. As discussed in Chapter 3, admixture has greatly impacted the genetic ancestry of the immune system, both adaptive and innate, of admixed Latin American populations in terms of specific disease- and health-related variants and regulation of gene expression. Any genomic approaches to healthcare in Latin America will require an understanding of the admixture profiles of the Latin American populations.

The process of admixture also influenced the ability of the Latin American populations to adapt to the environment, through admixture-enabled selection. The ancestral populations had previously been geographically isolated for tens of thousands of years and had genetically diverged from one another while adapting to their respective environments. Pre-adapted alleles with beneficial utility in the New World were stitched together in the admixed American populations such that they were better able to adapt to the environment in a shorter amount of time than what would be seen with new mutation. As discussed in Chapter 4, rapid admixture-enabled selection occurred on both single loci and polygenic phenotypes in the ~500 years since the Columbian Exchange began. There was strong selection on single haplotypes owing to their adaptive use in the environment and weaker selection on polygenic phenotypes.

As admixture acts to break down population structure, assortative mating acts to maintain it. Mate choice is a fundamental feature of human behavior and can maintain structure among distinct population groups even when they are co-located, such as those in Latin America. Latin American populations mate assortatively, with individuals choosing partners more like themselves in terms of genetic ancestry based on phenotypic cues. These phenotypic cues can be conscious or subconscious decisions, but ultimately they result in individuals choosing partners more like themselves than by random chance.

As discussed in Chapter 5, the development of a novel local ancestry approach to identify patterns of ancestry-based assortative mating yielded insights into the phenotypic cues that underlie this process as well as the specific ancestry components that drive assortative mating for different traits. The traits with evidence of ancestry-based assortative mating in Latin American populations varied in their appearance across ancestral populations, uncovering the genetic ancestry of the trait through patterns of ancestry homozygosity at trait-associated loci.

This thesis explores the implications of admixture and assortative mating in modern Latin American populations. These demographic processes, however, are not isolated to solely Latin America. As humans migrated out of Africa they consistently met with new populations. Admixture and assortative mating in these groups aided modern humans' ability to adapt to their environments and eventually migrate to other places around the world. By studying the impacts of these two processes in other admixed populations, we will continue to see how genetic ancestry influences many aspects of human evolution.

Further study into the relationship of genetic ancestry and health, specifically in admixed populations, will also help to identify ancestry-specific components that can lead to health disparities. Populations can be very genetically diverse even when co-located in the same country or even the same city; these same groups can also have different rates of health outcomes, including increased rates of disease when compared with their direct neighbors. This thesis has shown that there is a direct link between genetic ancestry and health- and disease-related phenotypes, but more work needs to be done to more deeply characterize this relationship. Having more detailed information and studies understanding the genetic ancestry architecture of health traits may assist in getting the right interventions

out to the populations who would most benefit from them and can help to guide healthcare workers' decisions when selecting treatments. Understanding this relationship can help to reduce the extent of health disparities seen between different admixed population groups.

APPENDIX A. SUPPLEMENTARY TABLES FOR CHAPTER 3

Table 5. Lists of all SNPs that show significant ancestry-enrichment.

For each admixed Latin population, all ancestry-enriched SNPs ($q < 0.05$) are given along with their meta-information, observed and expected reference and alternate allele frequencies, and significance values. The allele (reference or alternate) and enriched ancestry (African, European, or Native American) are designated for each ancestry-enriched SNP.

Chr[1]	Pos[2]	rsID[3]	Ref[4]	Alt[5]	Obs Ref Freq[6]	Obs Alt Freq[7]	Exp Ref Freq[8]	Exp Alt Freq[9]	χ^2 Statistic[10]	P-value [11]	*q*-value [12]	Anc Enriched Allele[13]	Enriched Anc[14]
Colombia													
1	11854476	rs1801131	T	G	0.87	0.13	0.75	0.25	12.69	3.67E-04	6.63E-03	Ref	AFR
1	11856378	rs1801133	G	A	0.46	0.54	0.57	0.43	10.53	1.17E-03	1.67E-02	Alt	NAT
1	12267292	rs3397	C	T	0.32	0.68	0.50	0.50	23.17	1.48E-06	6.50E-05	Alt	EUR
1	16505320	rs1497406	A	G	0.51	0.49	0.41	0.59	7.97	4.75E-03	4.60E-02	Ref	AFR
1	32756439	rs1741981	T	C	0.75	0.25	0.57	0.43	23.70	1.13E-06	5.12E-05	Ref	EUR
1	56112774	rs10443215	C	T	0.31	0.69	0.42	0.58	9.63	1.92E-03	2.47E-02	Alt	NAT
1	65992625	rs1751492	C	T	0.33	0.67	0.44	0.56	9.51	2.04E-03	2.61E-02	Alt	EUR
1	66058513	rs1137101	A	G	0.60	0.40	0.47	0.53	12.31	4.52E-04	7.68E-03	Ref	EUR
1	89849574	rs928655	G	A	0.29	0.71	0.41	0.59	12.62	3.81E-04	6.70E-03	Alt	EUR
1	96943994	rs1973993	T	C	0.48	0.52	0.38	0.62	8.17	4.26E-03	4.29E-02	Alt	NAT
1	114677948	rs529989	C	A	0.31	0.69	0.42	0.58	7.91	4.92E-03	4.75E-02	Alt	NAT
1	155279482	rs2297480	T	G	0.72	0.28	0.62	0.38	9.00	2.69E-03	3.01E-02	Ref	EUR
1	155868625	rs2282301	G	A	0.74	0.26	0.63	0.37	9.16	2.48E-03	2.95E-02	Ref	EUR
1	156879580	rs3737224	C	T	0.91	0.09	0.82	0.18	11.63	6.48E-04	1.02E-02	Ref	EUR
1	156881959	rs41273215	C	T	0.91	0.09	0.82	0.18	10.38	1.28E-03	1.77E-02	Ref	EUR
1	156883215	rs822442	C	A	0.91	0.09	0.82	0.18	10.38	1.28E-03	1.77E-02	Ref	EUR
1	196679455	rs10737680	A	C	0.43	0.57	0.56	0.44	13.52	2.36E-04	4.59E-03	Alt	AFR
1	201345487	rs12564445	G	A	0.84	0.16	0.75	0.25	8.09	4.46E-03	4.39E-02	Ref	EUR

Table 5 (continued).

1	206946407	rs1800872	T	G	0.30	0.70	0.42	0.58	10.61	1.13E-03	1.62E-02	Alt	EUR
1	206946634	rs1800871	A	G	0.30	0.70	0.42	0.58	10.61	1.13E-03	1.62E-02	Alt	EUR
1	226019653	rs1131873	G	A	0.88	0.12	0.78	0.22	10.11	1.48E-03	2.01E-02	Ref	AFR
1	230294916	rs2144300	C	T	0.42	0.58	0.54	0.46	10.36	1.29E-03	1.79E-02	Alt	EUR
1	232756026	rs10495332	T	C	0.98	0.02	0.90	0.10	14.99	1.08E-04	2.38E-03	Ref	AFR
1	237028564	rs3768142	G	T	0.30	0.70	0.42	0.58	9.66	1.88E-03	2.44E-02	Alt	AFR
1	247675559	rs7550918	C	T	0.30	0.70	0.14	0.86	38.92	4.41E-10	4.68E-08	Alt	NAT
2	10903412	rs1198872	C	T	0.55	0.45	0.65	0.35	9.41	2.16E-03	2.71E-02	Ref	EUR
2	21225281	rs1042034	C	T	0.19	0.81	0.34	0.66	18.57	1.64E-05	5.26E-04	Alt	AFR
2	21231524	rs676210	G	A	0.81	0.19	0.66	0.34	19.92	8.06E-06	2.86E-04	Ref	AFR
2	21288321	rs562338	A	G	0.25	0.75	0.16	0.84	9.89	1.66E-03	2.20E-02	Alt	NAT
2	25131316	rs6545814	A	G	0.63	0.37	0.50	0.50	12.26	4.64E-04	7.84E-03	Ref	NAT
2	56120853	rs1430193	A	T	0.64	0.36	0.45	0.55	27.89	1.28E-07	7.97E-06	Ref	EUR
2	111907691	rs724710	T	C	0.54	0.46	0.29	0.71	54.38	1.65E-13	2.89E-11	Ref	EUR
2	113598107	rs4848306	G	A	0.63	0.37	0.53	0.47	8.53	3.48E-03	3.69E-02	Ref	AFR
2	134266001	rs16826005	A	G	0.96	0.04	0.88	0.12	12.68	3.69E-04	6.65E-03	Ref	EUR
2	152981335	rs16830728	G	T	0.93	0.07	0.75	0.25	32.79	1.02E-08	8.05E-07	Ref	EUR
2	159899489	rs7582141	G	T	0.81	0.19	0.92	0.08	28.98	7.32E-08	4.76E-06	Ref	EUR
2	159899913	rs6432512	C	T	0.81	0.19	0.92	0.08	31.95	1.58E-08	1.18E-06	Ref	EUR
2	159936391	rs264588	C	A	0.81	0.19	0.93	0.07	40.00	2.54E-10	2.77E-08	Ref	EUR
2	159950865	rs264631	C	G	0.80	0.20	0.92	0.08	38.32	5.99E-10	6.23E-08	Ref	EUR
2	165513091	rs10195252	T	C	0.69	0.31	0.57	0.43	9.60	1.95E-03	2.51E-02	Ref	NAT
2	200638509	rs12615435	T	G	0.94	0.06	0.85	0.15	10.74	1.05E-03	1.52E-02	Ref	AFR
2	234668570	rs887829	C	T	0.66	0.34	0.75	0.25	8.20	4.19E-03	4.22E-02	Ref	NAT
3	12475557	rs3856806	C	T	0.95	0.05	0.86	0.14	11.43	7.24E-04	1.12E-02	Ref	AFR

Table 5 (continued).

3	26543420	rs2202157	C	T	0.95	0.05	0.85	0.15	12.50	4.07E-04	7.04E-03	Ref	AFR
3	119537291	rs3814058	T	C	0.84	0.16	0.71	0.29	14.97	1.10E-04	2.40E-03	Ref	EUR
3	119813282	rs334558	A	G	0.64	0.36	0.52	0.48	10.32	1.32E-03	1.82E-02	Ref	EUR
3	160429869	rs7624766	A	G	0.63	0.37	0.52	0.48	9.40	2.17E-03	2.72E-02	Ref	EUR
3	170725542	rs10513686	G	A	0.77	0.23	0.87	0.13	14.95	1.11E-04	2.42E-03	Ref	NAT
4	2906707	rs4961	G	T	0.84	0.16	0.74	0.26	8.09	4.46E-03	4.39E-02	Ref	AFR
4	3006043	rs1024323	C	T	0.55	0.45	0.66	0.34	8.49	3.57E-03	3.76E-02	Ref	NAT
4	6058497	rs16838131	G	A	0.98	0.02	1.00	0.00	9.05	2.63E-03	2.95E-02	Ref	NAT
4	9926967	rs13129697	T	G	0.45	0.55	0.64	0.36	31.75	1.76E-08	1.31E-06	Alt	AFR
4	38139024	rs9852	C	T	0.91	0.09	0.82	0.18	11.37	7.44E-04	1.15E-02	Ref	AFR
4	100163873	rs2201728	G	A	0.45	0.55	0.65	0.35	33.71	6.38E-09	5.18E-07	Alt	EUR
4	100239319	rs1229984	T	C	0.07	0.93	0.25	0.75	30.89	2.73E-08	1.95E-06	Alt	AFR
4	100495488	rs1800591	G	T	0.85	0.15	0.75	0.25	10.10	1.48E-03	2.01E-02	Ref	NAT
4	139493398	rs1450439	A	G	0.70	0.30	0.82	0.18	18.99	1.31E-05	4.35E-04	Ref	NAT
4	146875551	rs723794	G	T	0.29	0.71	0.41	0.59	10.65	1.10E-03	1.59E-02	Alt	EUR
4	154609523	rs1816702	T	C	0.15	0.85	0.09	0.91	7.83	5.15E-03	4.92E-02	Alt	NAT
4	159630817	rs8396	T	C	0.86	0.14	0.73	0.27	16.82	4.12E-05	1.14E-03	Ref	NAT
5	7870973	rs1801394	A	G	0.72	0.28	0.59	0.41	13.75	2.09E-04	4.15E-03	Ref	NAT
5	33951693	rs16891982	C	G	0.36	0.64	0.47	0.53	9.41	2.16E-03	2.71E-02	Alt	EUR
5	63250851	rs1364043	T	G	0.82	0.18	0.70	0.30	13.32	2.63E-04	5.01E-03	Ref	AFR
5	74642855	rs17244841	A	T	0.95	0.05	0.99	0.01	24.76	6.48E-07	3.09E-05	Ref	NAT
5	74655498	rs17238540	T	G	0.96	0.04	0.99	0.01	12.63	3.79E-04	6.66E-03	Ref	NAT
5	131819921	rs2070729	C	A	0.39	0.61	0.51	0.49	11.26	7.92E-04	1.21E-02	Alt	AFR
5	137707315	rs757647	G	A	0.74	0.26	0.63	0.37	10.04	1.53E-03	2.06E-02	Ref	EUR
5	167500460	rs13358864	T	A	0.95	0.05	0.86	0.14	12.50	4.07E-04	7.04E-03	Ref	AFR

Table 5 (continued).

5	176836532	rs1801020	A	G	0.25	0.75	0.38	0.62	14.07	1.76E-04	3.60E-03	Alt	EUR
6	7102084	rs675209	T	C	0.36	0.64	0.52	0.48	19.17	1.20E-05	4.02E-04	Alt	EUR
6	7801112	rs6923462	T	C	0.84	0.16	0.91	0.09	10.38	1.27E-03	1.77E-02	Ref	NAT
6	12296255	rs5370	G	T	0.85	0.15	0.76	0.24	8.44	3.66E-03	3.80E-02	Ref	AFR
6	31111356	rs9263739	C	T	0.87	0.13	0.78	0.22	7.99	4.72E-03	4.58E-02	Ref	NAT
6	31274380	rs9264942	T	C	0.70	0.30	0.60	0.40	8.83	2.96E-03	3.24E-02	Ref	AFR
6	32667910	rs2647044	G	A	0.96	0.04	0.89	0.11	8.06	4.53E-03	4.43E-02	Ref	NAT
6	32975014	rs399604	T	C	0.72	0.28	0.54	0.46	26.21	3.06E-07	1.74E-05	Ref	AFR
6	33283766	rs3130100	T	C	0.38	0.62	0.51	0.49	13.31	2.64E-04	5.03E-03	Alt	AFR
6	35369806	rs1883322	C	T	0.21	0.79	0.34	0.66	12.63	3.80E-04	6.68E-03	Alt	NAT
6	35395010	rs3734254	C	T	0.20	0.80	0.31	0.69	11.00	9.13E-04	1.36E-02	Alt	NAT
6	99782879	rs17059400	A	C	0.74	0.26	0.86	0.14	21.60	3.35E-06	1.33E-04	Ref	EUR
6	110777962	rs6907567	A	G	0.78	0.22	0.68	0.32	8.76	3.08E-03	3.35E-02	Ref	EUR
6	110778128	rs714368	T	C	0.78	0.22	0.68	0.32	7.93	4.86E-03	4.70E-02	Ref	EUR
6	121748542	rs11154022	A	G	0.28	0.72	0.40	0.60	11.73	6.13E-04	9.82E-03	Alt	AFR
6	122146034	rs9398652	C	A	0.89	0.11	0.79	0.21	11.68	6.32E-04	1.00E-02	Ref	EUR
6	135419018	rs9399137	T	C	0.88	0.12	0.76	0.24	15.23	9.54E-05	2.14E-03	Ref	AFR
6	137102365	rs9376230	C	A	0.30	0.70	0.41	0.59	8.80	3.01E-03	3.30E-02	Alt	AFR
6	154487421	rs2281617	C	T	0.88	0.12	0.73	0.27	22.63	1.97E-06	8.39E-05	Ref	EUR
6	155929801	rs35229355	C	T	0.98	0.02	0.85	0.15	26.23	3.03E-07	1.73E-05	Ref	EUR
7	29014195	rs2018683	G	T	0.46	0.54	0.57	0.43	8.68	3.22E-03	3.47E-02	Alt	EUR
7	80236014	rs13236689	T	G	0.65	0.35	0.54	0.46	10.36	1.29E-03	1.79E-02	Ref	EUR
7	102519031	rs17135875	T	C	0.73	0.27	0.85	0.15	17.98	2.23E-05	6.77E-04	Ref	NAT
8	11643915	rs804292	G	A	0.23	0.77	0.15	0.85	7.99	4.70E-03	4.57E-02	Alt	NAT
8	132330716	rs10108033	T	C	0.16	0.84	0.26	0.74	9.06	2.61E-03	2.95E-02	Alt	AFR

Table 5 (continued).

9	4744743	rs409801	T	C	0.65	0.35	0.53	0.47	12.29	4.55E-04	7.74E-03	Ref	AFR
9	4763176	rs385893	T	C	0.60	0.40	0.44	0.56	20.79	5.14E-06	1.94E-04	Ref	AFR
9	14446001	rs1556032	C	T	0.41	0.59	0.59	0.41	23.86	1.04E-06	4.73E-05	Alt	NAT
9	35141705	rs10972341	A	G	0.29	0.71	0.46	0.54	23.30	1.39E-06	6.14E-05	Alt	AFR
9	93636664	rs290227	G	A	0.81	0.19	0.66	0.34	19.78	8.71E-06	3.06E-04	Ref	EUR
9	104378003	rs10121600	C	T	0.62	0.38	0.72	0.28	9.49	2.07E-03	2.64E-02	Ref	NAT
9	108967088	rs2090409	C	A	0.73	0.27	0.59	0.41	14.87	1.15E-04	2.51E-03	Ref	AFR
9	124565820	rs10760187	T	C	0.60	0.40	0.37	0.63	42.08	8.74E-11	1.04E-08	Ref	EUR
9	132501881	rs2302821	A	C	0.85	0.15	0.76	0.24	8.32	3.93E-03	4.01E-02	Ref	EUR
10	23369421	rs722258	C	T	0.56	0.44	0.66	0.34	9.55	2.00E-03	2.57E-02	Ref	NAT
10	26734587	rs2992257	C	T	0.88	0.12	0.69	0.31	30.54	3.26E-08	2.30E-06	Ref	AFR
10	64963449	rs4379723	T	C	0.65	0.35	0.51	0.49	14.40	1.48E-04	3.10E-03	Ref	AFR
10	65104500	rs7896518	A	G	0.66	0.34	0.56	0.44	7.81	5.20E-03	4.96E-02	Ref	AFR
10	65133822	rs7923609	A	G	0.65	0.35	0.52	0.48	13.32	2.62E-04	5.01E-03	Ref	AFR
10	79211262	rs603788	G	C	0.40	0.60	0.54	0.46	13.37	2.55E-04	4.91E-03	Alt	EUR
10	96541616	rs4244285	G	A	0.89	0.11	0.80	0.20	10.69	1.08E-03	1.56E-02	Ref	EUR
10	96748495	rs1934969	A	T	0.55	0.45	0.43	0.57	12.53	4.00E-04	6.98E-03	Ref	NAT
10	96798548	rs1934951	C	T	0.87	0.13	0.71	0.29	23.13	1.51E-06	6.63E-05	Ref	EUR
10	101605693	rs3740065	A	G	0.91	0.09	0.81	0.19	12.67	3.71E-04	6.66E-03	Ref	EUR
11	12159661	rs1994318	C	A	0.52	0.48	0.42	0.58	7.88	4.99E-03	4.82E-02	Ref	AFR
11	13293905	rs900145	C	T	0.45	0.55	0.35	0.65	9.41	2.16E-03	2.71E-02	Alt	EUR
11	17409572	rs5219	T	C	0.20	0.80	0.35	0.65	20.87	4.91E-06	1.87E-04	Alt	AFR
11	27670108	rs10501087	T	C	0.81	0.19	0.71	0.29	9.28	2.32E-03	2.86E-02	Ref	AFR
11	27679916	rs6265	C	T	0.84	0.16	0.72	0.28	13.90	1.93E-04	3.86E-03	Ref	AFR
11	27684517	rs11030104	A	G	0.81	0.19	0.70	0.30	10.17	1.42E-03	1.95E-02	Ref	AFR

Table 5 (continued).

11	27700125	rs7103411	C	T	0.20	0.80	0.32	0.68	13.98	1.85E-04	3.75E-03	Alt	AFR
11	32895664	rs10767971	T	C	0.32	0.68	0.46	0.54	13.39	2.52E-04	4.87E-03	Alt	AFR
11	35123051	rs1559759	C	A	0.87	0.13	0.78	0.22	9.01	2.68E-03	2.99E-02	Ref	AFR
11	61557803	rs102275	T	C	0.49	0.51	0.65	0.35	18.18	2.01E-05	6.18E-04	Alt	AFR
11	69462910	rs9344	G	A	0.67	0.33	0.48	0.52	29.21	6.49E-08	4.30E-06	Ref	AFR
11	80377052	rs17140547	C	T	0.99	0.01	0.92	0.08	13.39	2.53E-04	4.87E-03	Ref	AFR
11	81235150	rs2032381	G	T	0.95	0.05	0.84	0.16	17.03	3.67E-05	1.03E-03	Ref	AFR
11	113306765	rs4436578	C	T	0.31	0.69	0.22	0.78	7.85	5.09E-03	4.89E-02	Alt	EUR
11	119099906	rs4938642	G	C	0.95	0.05	0.88	0.12	8.37	3.81E-03	3.92E-02	Ref	AFR
12	2757769	rs2239128	T	C	0.18	0.82	0.31	0.69	14.36	1.51E-04	3.15E-03	Alt	EUR
12	6451590	rs4149570	A	C	0.29	0.71	0.40	0.60	8.91	2.83E-03	3.13E-02	Alt	AFR
12	10170727	rs11053548	A	G	0.85	0.15	0.71	0.29	17.37	3.07E-05	8.87E-04	Ref	AFR
12	50350953	rs296766	T	C	0.02	0.98	0.07	0.93	9.34	2.24E-03	2.78E-02	Alt	EUR
12	51357542	rs12304921	A	G	0.85	0.15	0.73	0.27	14.23	1.61E-04	3.34E-03	Ref	AFR
12	54736470	rs4326844	A	G	0.25	0.75	0.42	0.58	22.36	2.26E-06	9.58E-05	Alt	AFR
12	58144665	rs2069502	C	T	0.79	0.21	0.64	0.36	20.61	5.64E-06	2.09E-04	Ref	AFR
12	77738005	rs6538140	G	A	0.30	0.70	0.40	0.60	8.83	2.96E-03	3.24E-02	Alt	AFR
12	88890671	rs995030	A	G	0.16	0.84	0.30	0.70	17.19	3.38E-05	9.58E-04	Alt	EUR
12	88953959	rs4474514	G	A	0.16	0.84	0.31	0.69	18.18	2.01E-05	6.19E-04	Alt	EUR
13	42505674	rs1900442	T	C	0.70	0.30	0.58	0.42	9.66	1.88E-03	2.44E-02	Ref	EUR
13	52566126	rs9535826	T	G	0.55	0.45	0.44	0.56	9.54	2.01E-03	2.58E-02	Ref	AFR
14	41523462	rs1959947	A	G	0.40	0.60	0.30	0.70	10.17	1.42E-03	1.95E-02	Alt	NAT
14	52781101	rs1353411	G	A	0.59	0.41	0.71	0.29	12.44	4.20E-04	7.22E-03	Ref	EUR
14	81598912	rs17111530	T	C	0.93	0.07	0.82	0.18	13.27	2.70E-04	5.11E-03	Ref	EUR
14	89594295	rs7158359	A	G	0.68	0.32	0.77	0.23	8.71	3.16E-03	3.42E-02	Ref	EUR

Table 5 (continued).

15	48392165	rs1834640	A	G	0.78	0.22	0.66	0.34	11.55	6.76E-04	1.06E-02	Ref	EUR
15	51545454	rs12907866	A	G	0.72	0.28	0.62	0.38	8.13	4.36E-03	4.38E-02	Ref	AFR
15	71424009	rs12904863	T	C	0.96	0.04	0.84	0.16	20.98	4.64E-06	1.78E-04	Ref	AFR
16	14388305	rs1659127	G	A	0.76	0.24	0.58	0.42	25.24	5.07E-07	2.68E-05	Ref	AFR
16	31110981	rs7196161	G	A	0.43	0.57	0.55	0.45	11.36	7.51E-04	1.16E-02	Alt	EUR
16	53876751	rs12595985	C	A	0.97	0.03	0.90	0.10	10.97	9.28E-04	1.37E-02	Ref	EUR
16	55844609	rs2244613	G	T	0.20	0.80	0.33	0.67	15.04	1.05E-04	2.32E-03	Alt	EUR
16	84046715	rs11864146	A	G	0.85	0.15	0.92	0.08	12.24	4.67E-04	7.86E-03	Ref	EUR
16	84987679	rs2326458	C	A	0.38	0.62	0.29	0.71	8.42	3.72E-03	3.83E-02	Alt	EUR
16	88713236	rs4673	A	G	0.39	0.61	0.30	0.70	8.24	4.10E-03	4.14E-02	Alt	NAT
16	89985940	rs2228479	G	A	0.99	0.01	0.91	0.09	15.73	7.31E-05	1.71E-03	Ref	AFR
16	89986608	rs2228478	A	G	0.93	0.07	0.84	0.16	11.46	7.10E-04	1.10E-02	Ref	EUR
17	19437187	rs2252281	T	C	0.71	0.29	0.59	0.41	10.65	1.10E-03	1.59E-02	Ref	NAT
18	5978931	rs1539808	C	T	0.97	0.03	0.90	0.10	11.48	7.05E-04	1.10E-02	Ref	EUR
18	55808073	rs520210	G	A	0.70	0.30	0.57	0.43	12.49	4.08E-04	7.05E-03	Ref	AFR
18	57673799	rs12964056	A	G	0.40	0.60	0.30	0.70	8.16	4.29E-03	4.31E-02	Alt	EUR
18	57851097	rs17782313	T	C	0.89	0.11	0.76	0.24	16.83	4.09E-05	1.13E-03	Ref	NAT
18	57884750	rs12970134	G	A	0.89	0.11	0.75	0.25	19.18	1.19E-05	4.01E-04	Ref	AFR
19	3595923	rs1131882	G	A	0.83	0.17	0.72	0.28	10.63	1.11E-03	1.60E-02	Ref	AFR
19	39735106	rs8103142	T	C	0.57	0.43	0.74	0.26	26.53	2.60E-07	1.50E-05	Ref	NAT
19	39738787	rs12979860	C	T	0.59	0.41	0.75	0.25	25.18	5.23E-07	2.68E-05	Ref	NAT
19	41515702	rs2279345	T	C	0.20	0.80	0.33	0.67	13.86	1.97E-04	3.93E-03	Alt	AFR
19	45403412	rs1160985	C	T	0.48	0.52	0.61	0.39	12.84	3.40E-04	6.20E-03	Alt	AFR
19	46172278	rs11671664	G	A	0.91	0.09	0.79	0.21	18.29	1.90E-05	5.99E-04	Ref	AFR
20	7106289	rs1884302	T	C	0.53	0.47	0.63	0.37	9.16	2.48E-03	2.95E-02	Ref	EUR

Table 5 (continued).

20	61981134	rs1044396	G	A	0.47	0.53	0.59	0.41	11.68	6.30E-04	1.00E-02	Alt	EUR
22	37469591	rs4820268	G	A	0.59	0.41	0.40	0.60	28.75	8.24E-08	5.29E-06	Ref	NAT
22	42526694	rs1065852	G	A	0.81	0.19	0.71	0.29	10.28	1.34E-03	1.84E-02	Ref	AFR
22	42528976	rs28360521	C	T	0.81	0.19	0.70	0.30	10.17	1.42E-03	1.95E-02	Ref	EUR
Mexico													
1	12267292	rs3397	C	T	0.31	0.69	0.54	0.46	28.37	1.00E-07	2.78E-06	Alt	EUR
1	16505320	rs1497406	A	G	0.54	0.46	0.33	0.67	23.67	1.14E-06	2.45E-05	Ref	AFR
1	32756439	rs1741981	T	C	0.76	0.24	0.56	0.44	21.38	3.77E-06	7.12E-05	Ref	EUR
1	48098406	rs2506991	A	G	0.34	0.66	0.49	0.51	11.28	7.82E-04	7.06E-03	Alt	EUR
1	55504650	rs2479409	G	A	0.66	0.34	0.50	0.50	13.78	2.05E-04	2.27E-03	Ref	NAT
1	56108604	rs1165472	A	G	0.91	0.09	0.80	0.20	10.86	9.83E-04	8.58E-03	Ref	NAT
1	56112774	rs10443215	C	T	0.21	0.79	0.34	0.66	10.01	1.56E-03	1.26E-02	Alt	NAT
1	65992625	rs1751492	C	T	0.32	0.68	0.56	0.44	30.51	3.33E-08	1.02E-06	Alt	EUR
1	66058513	rs1137101	A	G	0.52	0.48	0.37	0.63	13.45	2.45E-04	2.66E-03	Ref	EUR
1	70887099	rs672203	A	G	0.83	0.17	0.69	0.31	11.78	5.98E-04	5.64E-03	Ref	NAT
1	71514969	rs7551789	A	T	0.95	0.05	0.84	0.16	12.82	3.43E-04	3.58E-03	Ref	EUR
1	88132380	rs983332	G	T	0.87	0.13	0.73	0.27	12.74	3.58E-04	3.69E-03	Ref	EUR
1	98348885	rs1801265	G	A	0.24	0.76	0.15	0.85	8.90	2.85E-03	2.04E-02	Alt	NAT
1	109818530	rs646776	C	T	0.20	0.80	0.12	0.88	7.55	6.00E-03	3.75E-02	Alt	NAT
1	115837709	rs2239622	A	G	0.19	0.81	0.36	0.64	16.42	5.06E-05	6.85E-04	Alt	AFR
1	149892872	rs11205277	A	G	0.75	0.25	0.63	0.37	7.57	5.95E-03	3.74E-02	Ref	AFR
1	154418879	rs4537545	C	T	0.45	0.55	0.59	0.41	11.69	6.28E-04	5.89E-03	Alt	AFR
1	155279482	rs2297480	T	G	0.62	0.38	0.50	0.50	7.03	8.01E-03	4.74E-02	Ref	EUR
1	156869714	rs12041331	G	A	0.84	0.16	0.72	0.28	9.89	1.66E-03	1.33E-02	Ref	EUR
1	156873727	rs12407843	G	A	0.85	0.15	0.75	0.25	7.04	7.96E-03	4.74E-02	Ref	EUR

Table 5 (continued).

1	160743749	rs12068654	T	G	0.89	0.11	0.79	0.21	7.93	4.86E-03	3.13E-02	Ref	NAT
1	161915501	rs10918270	G	A	0.80	0.20	0.65	0.35	13.71	2.14E-04	2.35E-03	Ref	NAT
1	162085685	rs10494366	G	T	0.40	0.60	0.56	0.44	14.00	1.83E-04	2.07E-03	Alt	EUR
1	162112910	rs16857031	C	G	0.92	0.08	0.81	0.19	11.18	8.25E-04	7.40E-03	Ref	EUR
1	171091875	rs1795240	A	G	0.33	0.67	0.45	0.55	8.07	4.50E-03	2.96E-02	Alt	AFR
1	171254890	rs7877	C	T	0.56	0.44	0.73	0.27	17.34	3.12E-05	4.56E-04	Ref	NAT
1	177852580	rs633715	T	C	0.93	0.07	0.82	0.18	10.39	1.27E-03	1.07E-02	Ref	AFR
1	177913519	rs10913469	T	C	0.93	0.07	0.80	0.20	12.73	3.61E-04	3.71E-03	Ref	EUR
1	186947224	rs10157410	G	C	0.76	0.24	0.85	0.15	8.90	2.85E-03	2.04E-02	Ref	NAT
1	201345487	rs12564445	G	A	0.83	0.17	0.72	0.28	7.57	5.92E-03	3.72E-02	Ref	EUR
1	203135452	rs16851030	C	T	0.89	0.11	0.79	0.21	7.93	4.86E-03	3.13E-02	Ref	EUR
1	206946897	rs1800896	T	C	0.68	0.32	0.78	0.22	7.73	5.44E-03	3.47E-02	Ref	NAT
1	218931905	rs12037343	G	T	0.93	0.07	0.80	0.20	12.73	3.61E-04	3.71E-03	Ref	AFR
1	230294916	rs2144300	C	T	0.48	0.52	0.62	0.38	10.71	1.06E-03	9.18E-03	Alt	EUR
1	232756026	rs10495332	T	C	0.95	0.05	0.83	0.17	12.35	4.41E-04	4.36E-03	Ref	AFR
1	233719984	rs11800854	G	A	0.98	0.02	0.90	0.10	8.56	3.43E-03	2.38E-02	Ref	EUR
1	241894086	rs10926554	C	A	0.88	0.12	0.75	0.25	12.04	5.20E-04	5.00E-03	Ref	EUR
1	247675559	rs7550918	C	T	0.34	0.66	0.10	0.90	94.16	2.91E-22	4.86E-20	Alt	NAT
2	21225281	rs1042034	C	T	0.27	0.73	0.45	0.55	15.31	9.13E-05	1.13E-03	Alt	AFR
2	21231524	rs676210	G	A	0.73	0.27	0.55	0.45	15.31	9.13E-05	1.13E-03	Ref	AFR
2	25131316	rs6545814	A	G	0.71	0.29	0.54	0.46	15.22	9.58E-05	1.18E-03	Ref	NAT
2	27730940	rs1260326	T	C	0.35	0.65	0.50	0.50	11.28	7.83E-04	7.07E-03	Alt	AFR
2	27741237	rs780094	T	C	0.34	0.66	0.51	0.49	15.13	1.00E-04	1.22E-03	Alt	AFR
2	27742603	rs780093	T	C	0.34	0.66	0.51	0.49	15.13	1.00E-04	1.22E-03	Alt	AFR
2	31247514	rs9679162	G	T	0.37	0.63	0.55	0.45	16.68	4.43E-05	6.17E-04	Alt	AFR

Table 5 (continued).

2	31249427	rs12613732	T	G	0.44	0.56	0.64	0.36	22.94	1.67E-06	3.45E-05	Alt	NAT
2	56120853	rs1430193	A	T	0.62	0.38	0.34	0.66	42.42	7.35E-11	3.68E-09	Ref	EUR
2	111907691	rs724710	T	C	0.53	0.47	0.25	0.75	54.00	2.00E-13	1.46E-11	Ref	EUR
2	113594867	rs16944	A	G	0.53	0.47	0.40	0.60	9.42	2.15E-03	1.67E-02	Ref	AFR
2	128018063	rs3738948	A	G	0.88	0.12	0.71	0.29	18.40	1.79E-05	2.84E-04	Ref	AFR
2	152981335	rs16830728	G	T	0.94	0.06	0.63	0.37	51.14	8.60E-13	5.85E-11	Ref	EUR
2	159899489	rs7582141	G	T	0.84	0.16	0.92	0.08	8.06	4.54E-03	2.97E-02	Ref	EUR
2	159899913	rs6432512	C	T	0.84	0.16	0.91	0.09	8.06	4.54E-03	2.97E-02	Ref	EUR
2	159936391	rs264588	C	A	0.86	0.14	0.93	0.07	9.68	1.86E-03	1.47E-02	Ref	EUR
2	159950865	rs264631	C	G	0.86	0.14	0.92	0.08	6.94	8.42E-03	4.94E-02	Ref	EUR
2	165513091	rs10195252	T	C	0.83	0.17	0.66	0.34	15.44	8.50E-05	1.07E-03	Ref	NAT
2	166168503	rs2304016	A	G	0.99	0.01	0.93	0.07	8.79	3.03E-03	2.15E-02	Ref	AFR
2	166909544	rs3812718	C	T	0.66	0.34	0.53	0.47	9.07	2.60E-03	1.89E-02	Ref	AFR
2	170010985	rs2075252	T	C	0.22	0.78	0.39	0.61	15.89	6.73E-05	8.69E-04	Alt	AFR
2	174504924	rs13028485	G	A	0.94	0.06	0.84	0.16	8.53	3.49E-03	2.39E-02	Ref	EUR
2	206652300	rs10932125	C	G	0.69	0.31	0.56	0.44	8.13	4.36E-03	2.88E-02	Ref	AFR
2	216205167	rs16853826	G	A	0.80	0.20	0.69	0.31	7.23	7.18E-03	4.36E-02	Ref	EUR
2	234529643	rs6431558	C	T	0.59	0.41	0.47	0.53	8.03	4.60E-03	3.00E-02	Ref	EUR
2	234579892	rs3806598	A	C	0.95	0.05	0.86	0.14	7.82	5.16E-03	3.30E-02	Ref	EUR
2	234668570	rs887829	C	T	0.63	0.37	0.80	0.20	21.29	3.96E-06	7.46E-05	Ref	NAT
2	234669144	rs4148323	G	A	0.98	0.02	0.88	0.12	10.87	9.75E-04	8.53E-03	Ref	AFR
3	12267648	rs7616006	A	G	0.73	0.27	0.56	0.44	15.37	8.86E-05	1.10E-03	Ref	NAT
3	22473729	rs17011371	A	G	0.88	0.13	0.94	0.06	12.24	4.68E-04	4.54E-03	Ref	NAT
3	37574024	rs267567	G	A	0.27	0.73	0.42	0.58	10.43	1.24E-03	1.05E-02	Alt	NAT
3	38442490	rs2070488	G	A	0.74	0.26	0.62	0.38	8.46	3.62E-03	2.48E-02	Ref	AFR

Table 5 (continued).

3	45300605	rs33794	A	G	0.79	0.21	0.67	0.33	8.97	2.75E-03	1.98E-02	Ref	NAT
3	60001825	rs9311745	T	C	0.90	0.10	0.77	0.23	11.41	7.29E-04	6.67E-03	Ref	EUR
3	86916882	rs7642134	A	G	0.59	0.41	0.46	0.54	9.09	2.57E-03	1.88E-02	Ref	AFR
3	119525497	rs7643645	A	G	0.41	0.59	0.61	0.39	20.51	5.92E-06	1.05E-04	Alt	NAT
3	119537291	rs3814058	T	C	0.84	0.16	0.66	0.34	16.95	3.84E-05	5.48E-04	Ref	EUR
3	119631814	rs6438552	A	G	0.70	0.30	0.50	0.50	19.53	9.90E-06	1.67E-04	Ref	EUR
3	119813282	rs334558	A	G	0.73	0.27	0.49	0.51	30.04	4.24E-08	1.27E-06	Ref	EUR
3	132610752	rs6439371	G	A	0.52	0.48	0.30	0.70	31.48	2.02E-08	6.48E-07	Ref	AFR
3	152672779	rs6785504	G	T	0.38	0.62	0.54	0.46	12.58	3.91E-04	3.92E-03	Alt	AFR
3	156798473	rs1482853	C	A	0.65	0.35	0.53	0.47	7.06	7.89E-03	4.70E-02	Ref	AFR
3	170725542	rs10513686	G	A	0.80	0.20	0.91	0.09	15.54	8.08E-05	1.02E-03	Ref	NAT
3	184010048	rs3914188	G	C	0.20	0.80	0.32	0.68	10.24	1.37E-03	1.14E-02	Alt	EUR
3	187456709	rs3733017	T	G	0.98	0.02	0.88	0.12	12.76	3.54E-04	3.66E-03	Ref	EUR
4	2906707	rs4961	G	T	0.77	0.23	0.65	0.35	7.79	5.25E-03	3.35E-02	Ref	AFR
4	8503359	rs1949733	A	G	0.51	0.49	0.33	0.67	18.75	1.49E-05	2.41E-04	Ref	NAT
4	9926967	rs13129697	T	G	0.46	0.54	0.62	0.38	13.23	2.76E-04	2.94E-03	Alt	AFR
4	15964863	rs4698433	G	T	0.45	0.55	0.66	0.34	27.46	1.61E-07	4.28E-06	Alt	EUR
4	16893893	rs1483012	G	A	0.58	0.42	0.69	0.31	8.30	3.97E-03	2.69E-02	Ref	NAT
4	16908004	rs6819013	A	G	0.70	0.30	0.80	0.20	8.40	3.75E-03	2.55E-02	Ref	NAT
4	38139024	rs9852	C	T	0.89	0.11	0.76	0.24	12.30	4.52E-04	4.44E-03	Ref	AFR
4	88213808	rs6834314	A	G	0.87	0.13	0.71	0.29	15.21	9.64E-05	1.18E-03	Ref	AFR
4	100163873	rs2201728	G	A	0.44	0.56	0.71	0.29	43.27	4.78E-11	2.47E-09	Alt	EUR
4	100239319	rs1229984	T	C	0.09	0.91	0.39	0.61	49.92	1.60E-12	1.03E-10	Alt	AFR
4	139493398	rs1450439	A	G	0.71	0.29	0.86	0.14	23.34	1.36E-06	2.86E-05	Ref	NAT
4	146794621	rs4547811	T	C	0.44	0.56	0.71	0.29	46.57	8.84E-12	5.03E-10	Alt	NAT

Table 5 (continued).

4	146875551	rs723794	G	T	0.31	0.69	0.43	0.57	7.17	7.40E-03	4.48E-02	Alt	EUR
4	154609523	rs1816702	T	C	0.13	0.88	0.06	0.94	12.24	4.68E-04	4.54E-03	Alt	NAT
4	155514879	rs13109457	G	A	0.75	0.25	0.64	0.36	7.57	5.95E-03	3.74E-02	Ref	EUR
4	175071602	rs12507634	A	G	0.44	0.56	0.72	0.28	50.09	1.47E-12	9.52E-11	Alt	EUR
4	182197947	rs1454694	T	C	0.89	0.11	0.79	0.21	7.93	4.86E-03	3.13E-02	Ref	AFR
5	4831601	rs816475	T	C	0.73	0.27	0.83	0.17	7.90	4.93E-03	3.17E-02	Ref	NAT
5	8652870	rs200113	T	C	0.89	0.11	0.80	0.20	6.95	8.38E-03	4.94E-02	Ref	AFR
5	40679567	rs4133101	T	C	0.35	0.65	0.47	0.53	8.02	4.63E-03	3.01E-02	Alt	AFR
5	57214817	rs10041935	A	C	0.82	0.18	0.68	0.32	12.79	3.48E-04	3.61E-03	Ref	AFR
5	58713680	rs2547917	G	A	0.92	0.08	0.79	0.21	13.57	2.30E-04	2.52E-03	Ref	AFR
5	59736773	rs702553	A	T	0.73	0.27	0.53	0.47	21.21	4.12E-06	7.69E-05	Ref	EUR
5	63250851	rs1364043	T	G	0.78	0.22	0.62	0.38	14.58	1.34E-04	1.59E-03	Ref	AFR
5	63258565	rs6295	C	G	0.58	0.42	0.42	0.58	14.20	1.64E-04	1.90E-03	Ref	AFR
5	63261329	rs10042486	C	T	0.55	0.45	0.37	0.63	17.79	2.47E-05	3.70E-04	Ref	EUR
5	73276903	rs6894385	A	C	0.91	0.09	0.78	0.22	12.88	3.31E-04	3.46E-03	Ref	AFR
5	74642855	rs17244841	A	T	0.96	0.04	0.99	0.01	16.13	5.93E-05	7.76E-04	Ref	NAT
5	74655498	rs17238540	T	G	0.96	0.04	0.99	0.01	16.13	5.93E-05	7.76E-04	Ref	NAT
5	75514986	rs11960832	C	T	0.48	0.52	0.65	0.35	16.59	4.65E-05	6.41E-04	Alt	AFR
5	76781471	rs163030	A	C	0.27	0.73	0.44	0.56	15.37	8.86E-05	1.10E-03	Alt	AFR
5	88183651	rs17560407	A	G	0.88	0.13	0.77	0.23	7.53	6.05E-03	3.78E-02	Ref	AFR
5	93810208	rs6869388	T	C	0.79	0.21	0.88	0.12	10.87	9.75E-04	8.53E-03	Ref	NAT
5	137707315	rs757647	G	A	0.74	0.26	0.57	0.43	15.43	8.56E-05	1.07E-03	Ref	EUR
5	148206473	rs1042714	G	C	0.14	0.86	0.25	0.75	8.17	4.27E-03	2.85E-02	Alt	NAT
6	2235633	rs9378688	G	A	0.96	0.04	0.83	0.17	14.58	1.34E-04	1.59E-03	Ref	AFR
6	7102084	rs675209	T	C	0.48	0.52	0.63	0.37	13.45	2.45E-04	2.66E-03	Alt	EUR

Table 5 (continued).

6	7801112	rs6923462	T	C	0.87	0.13	0.94	0.06	10.80	1.02E-03	8.80E-03	Ref	NAT
6	18139802	rs12201199	A	T	0.89	0.11	0.95	0.05	11.19	8.22E-04	7.39E-03	Ref	NAT
6	26233387	rs10946808	A	G	0.69	0.31	0.48	0.52	21.15	4.26E-06	7.87E-05	Ref	AFR
6	31093587	rs3815087	G	A	0.81	0.19	0.70	0.30	7.34	6.76E-03	4.13E-02	Ref	EUR
6	31253444	rs9461684	C	T	0.80	0.20	0.89	0.11	9.70	1.84E-03	1.46E-02	Ref	NAT
6	31431691	rs2255221	G	T	0.87	0.13	0.93	0.07	7.65	5.68E-03	3.59E-02	Ref	NAT
6	32975014	rs399604	T	C	0.69	0.31	0.51	0.49	15.14	9.98E-05	1.22E-03	Ref	AFR
6	33055538	rs9277554	C	T	0.77	0.23	0.54	0.46	26.44	2.71E-07	6.91E-06	Ref	EUR
6	33283766	rs3130100	T	C	0.42	0.58	0.56	0.44	9.14	2.50E-03	1.88E-02	Alt	AFR
6	35369806	rs1883322	C	T	0.12	0.88	0.31	0.69	22.73	1.87E-06	3.80E-05	Alt	NAT
6	35395010	rs3734254	C	T	0.12	0.88	0.29	0.71	18.40	1.79E-05	2.84E-04	Alt	NAT
6	35402785	rs4713858	A	G	0.06	0.94	0.20	0.80	15.64	7.67E-05	9.73E-04	Alt	AFR
6	39325078	rs20455	A	G	0.68	0.32	0.54	0.46	10.19	1.41E-03	1.17E-02	Ref	EUR
6	45095163	rs9395066	A	C	0.67	0.33	0.45	0.55	26.60	2.50E-07	6.43E-06	Ref	AFR
6	73749861	rs9351963	A	C	0.67	0.33	0.78	0.22	8.96	2.76E-03	1.98E-02	Ref	EUR
6	110777962	rs6907567	A	G	0.74	0.26	0.63	0.37	7.50	6.17E-03	3.85E-02	Ref	EUR
6	121748542	rs11154022	A	G	0.23	0.77	0.44	0.56	24.80	6.37E-07	1.44E-05	Alt	AFR
6	122146034	rs9398652	C	A	0.91	0.09	0.74	0.26	18.01	2.20E-05	3.34E-04	Ref	EUR
6	126964510	rs4273712	A	G	0.47	0.53	0.62	0.38	11.94	5.50E-04	5.23E-03	Alt	NAT
6	130008445	rs12660691	A	C	0.73	0.27	0.84	0.16	9.63	1.92E-03	1.51E-02	Ref	EUR
6	135419018	rs9399137	T	C	0.86	0.14	0.74	0.26	10.25	1.36E-03	1.14E-02	Ref	AFR
6	137673302	rs6928289	G	A	0.54	0.46	0.42	0.58	7.21	7.26E-03	4.41E-02	Ref	EUR
6	142767633	rs3748069	A	G	0.76	0.24	0.63	0.37	8.61	3.35E-03	2.34E-02	Ref	NAT
6	147680359	rs9390459	A	G	0.44	0.56	0.60	0.40	14.37	1.50E-04	1.75E-03	Alt	AFR
6	154487421	rs2281617	C	T	0.90	0.10	0.68	0.32	29.80	4.78E-08	1.42E-06	Ref	EUR

Table 5 (continued).

6	155929801	rs35229355	C	T	0.95	0.05	0.76	0.24	25.08	5.51E-07	1.27E-05	Ref	EUR
6	160551204	rs683369	G	C	0.05	0.95	0.17	0.83	14.05	1.78E-04	2.02E-03	Alt	AFR
6	160560845	rs628031	A	G	0.12	0.88	0.32	0.68	24.26	8.43E-07	1.85E-05	Alt	NAT
6	166579270	rs2305089	C	T	0.45	0.55	0.58	0.42	8.20	4.19E-03	2.81E-02	Alt	EUR
7	17284577	rs4410790	T	C	0.63	0.37	0.51	0.49	8.00	4.67E-03	3.03E-02	Ref	NAT
7	28004198	rs4722750	C	T	0.96	0.04	0.86	0.14	9.77	1.78E-03	1.42E-02	Ref	EUR
7	30699972	rs2270007	G	C	0.11	0.89	0.30	0.70	23.05	1.58E-06	3.28E-05	Alt	AFR
7	33060946	rs3750117	A	G	0.30	0.70	0.42	0.58	8.20	4.19E-03	2.81E-02	Alt	AFR
7	80236014	rs13236689	T	G	0.67	0.33	0.49	0.51	16.54	4.78E-05	6.55E-04	Ref	EUR
7	99361466	rs2242480	C	T	0.61	0.39	0.79	0.21	24.83	6.26E-07	1.42E-05	Ref	EUR
7	101809851	rs365836	A	G	0.70	0.30	0.82	0.18	13.57	2.30E-04	2.52E-03	Ref	NAT
7	127164958	rs6467136	A	G	0.54	0.46	0.35	0.65	21.65	3.28E-06	6.29E-05	Ref	EUR
7	154509324	rs12666280	T	C	0.45	0.55	0.56	0.44	7.14	7.53E-03	4.52E-02	Alt	EUR
8	1244224	rs17669535	C	G	0.96	0.04	0.88	0.12	7.55	6.00E-03	3.75E-02	Ref	AFR
8	2740502	rs641525	T	G	0.91	0.09	0.77	0.23	14.11	1.73E-04	1.97E-03	Ref	EUR
8	4078353	rs2407314	G	C	0.46	0.54	0.58	0.42	7.21	7.26E-03	4.41E-02	Alt	AFR
8	23059324	rs20575	C	G	0.40	0.60	0.27	0.73	10.07	1.51E-03	1.23E-02	Alt	NAT
8	69389217	rs1517114	C	G	0.32	0.68	0.21	0.79	9.20	2.42E-03	1.84E-02	Alt	NAT
8	118184783	rs13266634	C	T	0.76	0.24	0.63	0.37	8.61	3.35E-03	2.34E-02	Ref	AFR
8	122275906	rs7834765	G	T	0.45	0.55	0.62	0.38	16.00	6.32E-05	8.18E-04	Alt	NAT
8	126486409	rs17321515	A	G	0.63	0.38	0.49	0.51	10.13	1.45E-03	1.20E-02	Ref	EUR
8	126490972	rs2954029	A	T	0.64	0.36	0.50	0.50	9.03	2.65E-03	1.91E-02	Ref	AFR
8	129072161	rs2648875	G	A	0.45	0.55	0.60	0.40	11.77	6.03E-04	5.68E-03	Alt	NAT
8	139884509	rs6988229	C	T	0.82	0.18	0.92	0.08	18.33	1.86E-05	2.93E-04	Ref	NAT
9	4744743	rs409801	T	C	0.66	0.34	0.47	0.53	19.61	9.51E-06	1.61E-04	Ref	AFR

Table 5 (continued).

9	4763176	rs385893	T	C	0.61	0.39	0.35	0.65	37.32	1.00E-09	4.21E-08	Ref	AFR
9	4814948	rs13300663	G	C	0.86	0.14	0.69	0.31	17.60	2.73E-05	4.05E-04	Ref	AFR
9	27536397	rs2814707	C	T	0.95	0.05	0.86	0.14	7.82	5.16E-03	3.30E-02	Ref	NAT
9	27543281	rs3849942	T	C	0.05	0.95	0.15	0.85	8.90	2.85E-03	2.04E-02	Alt	NAT
9	35141705	rs10972341	A	G	0.27	0.73	0.55	0.45	40.86	1.64E-10	7.73E-09	Alt	AFR
9	35648950	rs3138083	A	G	0.52	0.48	0.35	0.65	15.11	1.01E-04	1.23E-03	Ref	AFR
9	86920236	rs7867504	T	C	0.31	0.69	0.49	0.51	15.14	9.98E-05	1.22E-03	Alt	AFR
9	93636664	rs290227	G	A	0.74	0.26	0.57	0.43	15.43	8.56E-05	1.07E-03	Ref	EUR
9	104223233	rs10819937	C	G	0.51	0.49	0.29	0.71	29.80	4.78E-08	1.42E-06	Ref	NAT
9	111455575	rs10512385	A	G	0.96	0.04	0.88	0.12	8.64	3.28E-03	2.30E-02	Ref	NAT
9	112521126	rs4978848	G	C	0.27	0.73	0.41	0.59	10.49	1.20E-03	1.02E-02	Alt	NAT
9	114293634	rs10980926	A	G	0.63	0.37	0.49	0.51	10.13	1.46E-03	1.20E-02	Ref	AFR
9	114301585	rs10441737	C	T	0.63	0.37	0.48	0.52	11.29	7.78E-04	7.04E-03	Ref	NAT
9	119249339	rs7852872	C	G	0.48	0.52	0.60	0.40	8.34	3.87E-03	2.63E-02	Alt	NAT
9	124565820	rs10760187	T	C	0.59	0.41	0.33	0.67	40.97	1.55E-10	7.34E-09	Ref	EUR
9	125137695	rs10306135	A	T	0.85	0.15	0.94	0.06	16.13	5.90E-05	7.76E-04	Ref	NAT
9	132501881	rs2302821	A	C	0.89	0.11	0.66	0.34	31.17	2.37E-08	7.48E-07	Ref	EUR
10	26734587	rs2992257	C	T	0.83	0.17	0.62	0.38	22.53	2.07E-06	4.16E-05	Ref	AFR
10	30834632	rs11008099	G	A	0.96	0.04	0.84	0.16	13.33	2.61E-04	2.81E-03	Ref	AFR
10	52010708	rs10508921	C	T	0.91	0.09	0.78	0.22	13.21	2.78E-04	2.96E-03	Ref	EUR
10	64963449	rs4379723	T	C	0.70	0.30	0.55	0.45	10.25	1.37E-03	1.14E-02	Ref	AFR
10	65027610	rs10761731	A	T	0.72	0.28	0.60	0.40	7.33	6.77E-03	4.13E-02	Ref	AFR
10	65104500	rs7896518	A	G	0.72	0.28	0.60	0.40	7.33	6.77E-03	4.13E-02	Ref	AFR
10	65133822	rs7923609	A	G	0.70	0.30	0.57	0.43	8.16	4.28E-03	2.86E-02	Ref	AFR
10	90826779	rs1937332	A	G	0.52	0.48	0.67	0.33	12.79	3.48E-04	3.61E-03	Ref	NAT

Table 5 (continued).

10	94839642	rs2068888	G	A	0.54	0.46	0.35	0.65	19.74	8.87E-06	1.52E-04	Ref	AFR
10	96541616	rs4244285	G	A	0.88	0.13	0.76	0.24	9.58	1.97E-03	1.54E-02	Ref	EUR
10	96748495	rs1934969	A	T	0.63	0.38	0.47	0.53	12.55	3.96E-04	3.97E-03	Ref	NAT
10	96798548	rs1934951	C	T	0.84	0.16	0.68	0.32	15.83	6.95E-05	8.93E-04	Ref	EUR
10	101605693	rs3740065	A	G	0.90	0.10	0.78	0.22	10.29	1.34E-03	1.12E-02	Ref	EUR
10	101795361	rs10883437	T	A	0.47	0.53	0.64	0.36	18.13	2.06E-05	3.18E-04	Alt	AFR
10	112836503	rs1800544	G	C	0.34	0.66	0.48	0.52	10.15	1.45E-03	1.19E-02	Alt	EUR
11	243268	rs505404	T	G	0.73	0.27	0.89	0.11	35.37	2.73E-09	1.04E-07	Ref	NAT
11	12159661	rs1994318	C	A	0.65	0.35	0.45	0.55	21.38	3.77E-06	7.12E-05	Ref	AFR
11	20659757	rs2298826	G	A	0.76	0.24	0.60	0.40	13.04	3.05E-04	3.21E-03	Ref	AFR
11	27670108	rs10501087	T	C	0.80	0.20	0.65	0.35	12.37	4.36E-04	4.33E-03	Ref	AFR
11	27679916	rs6265	C	T	0.80	0.20	0.65	0.35	13.71	2.14E-04	2.35E-03	Ref	AFR
11	27684517	rs11030104	A	G	0.80	0.20	0.63	0.37	16.27	5.48E-05	7.40E-04	Ref	AFR
11	27700125	rs7103411	C	T	0.20	0.80	0.38	0.62	19.05	1.28E-05	2.09E-04	Alt	AFR
11	32895664	rs10767971	T	C	0.32	0.68	0.49	0.51	15.13	1.00E-04	1.22E-03	Alt	AFR
11	35123051	rs1559759	C	A	0.91	0.09	0.75	0.25	19.76	8.77E-06	1.50E-04	Ref	AFR
11	61557803	rs102275	T	C	0.29	0.71	0.65	0.35	72.52	1.66E-17	1.84E-15	Alt	AFR
11	64048912	rs477895	C	T	0.13	0.87	0.23	0.77	7.36	6.68E-03	4.09E-02	Alt	EUR
11	69462910	rs9344	G	A	0.67	0.33	0.46	0.54	22.92	1.69E-06	3.47E-05	Ref	AFR
11	81235150	rs2032381	G	T	0.93	0.07	0.78	0.22	16.50	4.86E-05	6.60E-04	Ref	AFR
11	112026156	rs5744247	G	C	0.90	0.10	0.71	0.29	21.90	2.88E-06	5.62E-05	Ref	AFR
11	113280274	rs2734842	G	C	0.31	0.69	0.43	0.57	7.17	7.40E-03	4.48E-02	Alt	EUR
11	113281776	rs2734841	A	C	0.29	0.71	0.42	0.58	9.26	2.35E-03	1.79E-02	Alt	EUR
11	113282295	rs1124493	T	G	0.30	0.70	0.42	0.58	8.20	4.19E-03	2.81E-02	Alt	EUR
11	113283477	rs6275	A	G	0.29	0.71	0.42	0.58	9.26	2.35E-03	1.79E-02	Alt	EUR

Table 5 (continued).

11	116652207	rs12286037	C	T	0.90	0.10	0.96	0.04	8.57	3.42E-03	2.38E-02	Ref	NAT
11	131807171	rs12098973	A	G	0.88	0.12	0.77	0.23	8.74	3.12E-03	2.20E-02	Ref	EUR
12	988558	rs880054	C	T	0.45	0.55	0.33	0.67	9.07	2.60E-03	1.88E-02	Alt	NAT
12	2757769	rs2239128	T	C	0.11	0.89	0.33	0.67	27.78	1.36E-07	3.69E-06	Alt	EUR
12	6291093	rs7342306	G	A	0.77	0.23	0.64	0.36	10.89	9.65E-04	8.47E-03	Ref	AFR
12	10170727	rs11053548	A	G	0.91	0.09	0.73	0.27	20.80	5.09E-06	9.22E-05	Ref	AFR
12	11547532	rs2900174	A	G	0.84	0.16	0.92	0.08	10.85	9.89E-04	8.62E-03	Ref	EUR
12	11855624	rs2187642	A	C	0.28	0.72	0.49	0.51	22.79	1.81E-06	3.70E-05	Alt	EUR
12	20860093	rs3794271	G	A	0.56	0.44	0.31	0.69	40.16	2.34E-10	1.08E-08	Ref	AFR
12	21283322	rs4149015	G	A	0.98	0.02	0.90	0.10	8.56	3.43E-03	2.38E-02	Ref	AFR
12	21327740	rs4149036	C	A	0.88	0.13	0.64	0.36	32.31	1.31E-08	4.36E-07	Ref	EUR
12	21329738	rs2306283	A	G	0.63	0.38	0.39	0.61	29.54	5.48E-08	1.60E-06	Ref	EUR
12	21377559	rs4149080	G	C	0.90	0.10	0.70	0.30	23.39	1.32E-06	2.80E-05	Ref	EUR
12	21378021	rs4149081	G	A	0.90	0.10	0.70	0.30	23.39	1.32E-06	2.80E-05	Ref	EUR
12	51357542	rs12304921	A	G	0.87	0.13	0.67	0.33	22.15	2.52E-06	5.01E-05	Ref	AFR
12	54736470	rs4326844	A	G	0.18	0.82	0.44	0.56	36.56	1.48E-09	6.03E-08	Alt	AFR
12	58144665	rs2069502	C	T	0.70	0.30	0.55	0.45	11.38	7.42E-04	6.78E-03	Ref	AFR
12	65534624	rs6581612	C	A	0.27	0.73	0.13	0.87	21.98	2.76E-06	5.43E-05	Alt	NAT
12	88890671	rs995030	A	G	0.13	0.88	0.29	0.71	16.77	4.23E-05	5.95E-04	Alt	EUR
12	88953959	rs4474514	G	A	0.13	0.88	0.30	0.70	18.11	2.08E-05	3.20E-04	Alt	EUR
12	112190438	rs6490294	C	A	0.63	0.37	0.44	0.56	19.84	8.41E-06	1.45E-04	Ref	EUR
12	112817783	rs11066280	T	A	0.99	0.01	0.90	0.10	11.13	8.51E-04	7.55E-03	Ref	AFR
12	117327592	rs7294919	T	C	0.90	0.10	0.80	0.20	7.16	7.46E-03	4.50E-02	Ref	EUR
13	47469940	rs6313	G	A	0.67	0.33	0.51	0.49	12.51	4.04E-04	4.04E-03	Ref	AFR
13	47471478	rs6311	C	T	0.67	0.33	0.51	0.49	12.51	4.04E-04	4.04E-03	Ref	AFR

Table 5 (continued).

13	52566126	rs9535826	T	G	0.59	0.41	0.44	0.56	11.42	7.27E-04	6.67E-03	Ref	AFR
14	23977010	rs223116	A	G	0.45	0.55	0.33	0.67	7.88	5.00E-03	3.21E-02	Alt	EUR
14	48015982	rs1160351	A	C	0.80	0.20	0.48	0.52	50.05	1.50E-12	9.70E-11	Ref	AFR
14	81598912	rs17111530	T	C	0.94	0.06	0.79	0.21	15.64	7.67E-05	9.73E-04	Ref	EUR
14	92427222	rs7153027	A	C	0.79	0.21	0.67	0.33	7.97	4.75E-03	3.07E-02	Ref	NAT
14	105263608	rs2494752	A	G	0.27	0.73	0.44	0.56	16.73	4.31E-05	6.02E-04	Alt	EUR
15	29731444	rs11856574	G	A	0.84	0.16	0.73	0.27	8.85	2.93E-03	2.09E-02	Ref	EUR
15	48392165	rs1834640	A	G	0.66	0.34	0.51	0.49	11.28	7.82E-04	7.06E-03	Ref	EUR
15	51545454	rs12907866	A	G	0.73	0.27	0.58	0.42	11.56	6.73E-04	6.25E-03	Ref	AFR
15	51631279	rs7176005	C	T	0.88	0.12	0.78	0.22	7.73	5.44E-03	3.47E-02	Ref	EUR
15	71424009	rs12904863	T	C	0.95	0.05	0.78	0.22	20.16	7.12E-06	1.25E-04	Ref	AFR
15	100786271	rs4533267	A	G	0.38	0.62	0.23	0.77	17.83	2.41E-05	3.62E-04	Alt	NAT
16	14388305	rs1659127	G	A	0.77	0.23	0.55	0.45	23.06	1.57E-06	3.27E-05	Ref	AFR
16	23634026	rs420259	A	G	0.77	0.23	0.65	0.35	7.71	5.49E-03	3.49E-02	Ref	EUR
16	27375787	rs8832	A	G	0.40	0.60	0.52	0.48	7.04	7.98E-03	4.74E-02	Alt	EUR
16	30918487	rs11649653	C	G	0.54	0.46	0.34	0.66	23.67	1.14E-06	2.45E-05	Ref	AFR
16	31048079	rs10871454	C	T	0.52	0.48	0.34	0.66	18.32	1.87E-05	2.94E-04	Ref	AFR
16	31102321	rs7294	C	T	0.65	0.35	0.80	0.20	17.42	2.99E-05	4.42E-04	Ref	NAT
16	31103796	rs2359612	A	G	0.47	0.53	0.66	0.34	19.95	7.96E-06	1.38E-04	Alt	AFR
16	31104509	rs8050894	C	G	0.49	0.51	0.33	0.67	15.63	7.71E-05	9.77E-04	Alt	NAT
16	31104878	rs9934438	G	A	0.53	0.47	0.35	0.65	19.95	7.96E-06	1.38E-04	Ref	AFR
16	31107689	rs9923231	C	T	0.53	0.47	0.35	0.65	19.95	7.96E-06	1.38E-04	Ref	AFR
16	31110981	rs7196161	G	A	0.48	0.52	0.67	0.33	20.41	6.24E-06	1.11E-04	Alt	EUR
16	53876751	rs12595985	C	A	0.92	0.08	0.83	0.17	7.90	4.93E-03	3.17E-02	Ref	EUR
16	55793695	rs3785161	A	C	0.87	0.13	0.77	0.23	7.36	6.68E-03	4.09E-02	Ref	AFR

Table 5 (continued).

16	56994894	rs4783961	G	A	0.48	0.52	0.67	0.33	20.41	6.24E-06	1.11E-04	Alt	EUR
16	75167579	rs7195303	G	A	0.05	0.95	0.15	0.85	11.61	6.54E-04	6.11E-03	Alt	AFR
16	84046715	rs11864146	A	G	0.82	0.18	0.93	0.07	23.42	1.30E-06	2.76E-05	Ref	EUR
16	85961562	rs17444745	G	A	0.92	0.08	0.83	0.17	7.90	4.93E-03	3.17E-02	Ref	AFR
16	89985940	rs2228479	G	A	0.95	0.05	0.87	0.13	8.21	4.17E-03	2.81E-02	Ref	AFR
17	18232096	rs1979277	G	A	0.69	0.31	0.82	0.18	15.32	9.09E-05	1.13E-03	Ref	NAT
17	19437187	rs2252281	T	C	0.77	0.23	0.63	0.37	9.72	1.83E-03	1.45E-02	Ref	NAT
17	19804247	rs397969	T	C	0.59	0.41	0.72	0.28	12.74	3.58E-04	3.69E-03	Ref	NAT
17	48712087	rs4793665	C	T	0.38	0.62	0.27	0.73	7.71	5.50E-03	3.50E-02	Alt	NAT
17	48768486	rs1051640	A	G	0.96	0.04	0.88	0.12	7.55	6.00E-03	3.75E-02	Ref	NAT
17	49960309	rs967676	T	C	0.77	0.23	0.65	0.35	7.71	5.49E-03	3.49E-02	Ref	AFR
17	78324259	rs12051723	G	T	0.95	0.05	0.81	0.19	16.11	5.99E-05	7.83E-04	Ref	AFR
18	5978931	rs1539808	C	T	0.98	0.02	0.86	0.14	14.55	1.37E-04	1.61E-03	Ref	EUR
18	29038123	rs1941184	A	C	0.65	0.35	0.75	0.25	7.04	7.96E-03	4.74E-02	Ref	NAT
18	55816791	rs4149601	G	A	0.88	0.12	0.69	0.31	22.73	1.87E-06	3.80E-05	Ref	NAT
19	10000322	rs1862471	C	G	0.43	0.57	0.61	0.39	17.36	3.09E-05	4.52E-04	Alt	EUR
19	10395683	rs5498	A	G	0.38	0.63	0.64	0.36	39.23	3.77E-10	1.69E-08	Alt	EUR
19	10397403	rs3093030	C	T	0.38	0.63	0.67	0.33	51.17	8.46E-13	5.76E-11	Alt	EUR
19	29736342	rs11083866	G	A	0.74	0.26	0.60	0.40	10.56	1.16E-03	9.89E-03	Ref	AFR
19	39735106	rs8103142	T	C	0.54	0.46	0.81	0.19	57.46	3.44E-14	2.68E-12	Ref	NAT
19	39738787	rs12979860	C	T	0.54	0.46	0.81	0.19	62.82	2.26E-15	2.03E-13	Ref	NAT
19	41515702	rs2279345	T	C	0.21	0.79	0.33	0.67	7.97	4.75E-03	3.07E-02	Alt	AFR
19	45403412	rs1160985	C	T	0.49	0.51	0.63	0.37	10.89	9.65E-04	8.47E-03	Alt	AFR
19	45923653	rs11615	A	G	0.26	0.74	0.42	0.58	12.88	3.32E-04	3.47E-03	Alt	AFR
19	46172278	rs11671664	G	A	0.93	0.07	0.70	0.30	31.48	2.02E-08	6.48E-07	Ref	AFR

Table 5 (continued).

19	49206172	rs516246	C	T	0.68	0.32	0.78	0.22	7.73	5.44E-03	3.47E-02	Ref	NAT
19	49206985	rs602662	G	A	0.66	0.34	0.77	0.23	7.36	6.68E-03	4.09E-02	Ref	NAT
20	17122593	rs852069	A	G	0.44	0.56	0.59	0.41	11.62	6.51E-04	6.09E-03	Alt	EUR
20	45288453	rs6066043	G	A	0.76	0.24	0.62	0.38	9.63	1.91E-03	1.51E-02	Ref	AFR
20	60791404	rs3787429	C	T	0.44	0.56	0.60	0.40	14.37	1.50E-04	1.75E-03	Alt	AFR
20	61981104	rs1044397	C	T	0.43	0.57	0.58	0.42	11.56	6.73E-04	6.25E-03	Alt	EUR
20	61981134	rs1044396	G	A	0.44	0.56	0.65	0.35	24.98	5.78E-07	1.33E-05	Alt	EUR
21	28217320	rs402007	C	G	0.60	0.40	0.42	0.58	16.94	3.85E-05	5.48E-04	Ref	NAT
22	19951207	rs4818	C	G	0.76	0.24	0.62	0.38	10.71	1.06E-03	9.18E-03	Ref	AFR
22	19952132	rs4646316	C	T	0.84	0.16	0.69	0.31	13.13	2.91E-04	3.07E-03	Ref	AFR
22	42526694	rs1065852	G	A	0.85	0.15	0.61	0.39	31.54	1.95E-08	6.30E-07	Ref	AFR
22	42528976	rs28360521	C	T	0.85	0.15	0.61	0.39	31.54	1.95E-08	6.30E-07	Ref	EUR
22	44324727	rs738409	C	G	0.45	0.55	0.68	0.32	32.30	1.32E-08	4.39E-07	Alt	NAT
22	44333694	rs2896019	T	G	0.49	0.51	0.70	0.30	24.93	5.95E-07	1.37E-05	Alt	NAT
Peru													
1	11046855	rs9430161	G	T	0.94	0.06	0.87	0.13	6.32	1.20E-02	3.96E-02	Ref	NAT
1	11854476	rs1801131	T	G	0.92	0.08	0.77	0.23	20.80	5.11E-06	3.12E-05	Ref	AFR
1	12267292	rs3397	C	T	0.16	0.84	0.62	0.38	147.68	5.57E-34	2.60E-32	Alt	EUR
1	16505320	rs1497406	A	G	0.74	0.26	0.26	0.74	201.19	1.15E-45	7.90E-44	Ref	AFR
1	29174946	rs529520	A	C	0.45	0.55	0.24	0.76	44.76	2.23E-11	2.62E-10	Alt	NAT
1	29175373	rs581111	A	G	0.07	0.93	0.15	0.85	7.93	4.87E-03	1.75E-02	Alt	NAT
1	29190138	rs4654327	G	A	0.20	0.80	0.29	0.71	7.25	7.08E-03	2.45E-02	Alt	NAT
1	32756439	rs1741981	T	C	0.89	0.11	0.51	0.49	96.43	9.25E-23	2.54E-21	Ref	EUR
1	40433771	rs3103778	A	G	0.34	0.66	0.44	0.56	6.90	8.64E-03	2.95E-02	Alt	NAT
1	48098406	rs2506991	A	G	0.38	0.62	0.52	0.48	12.46	4.15E-04	1.86E-03	Alt	EUR

Table 5 (continued).

1	55504650	rs2479409	G	A	0.79	0.21	0.62	0.38	19.65	9.32E-06	5.49E-05	Ref	NAT
1	56108604	rs1165472	A	G	0.93	0.07	0.84	0.16	9.91	1.65E-03	6.61E-03	Ref	NAT
1	56112774	rs10443215	C	T	0.15	0.85	0.25	0.75	8.10	4.44E-03	1.60E-02	Alt	NAT
1	63049593	rs1748195	C	G	0.45	0.55	0.73	0.27	68.67	1.17E-16	2.17E-15	Alt	AFR
1	65992625	rs1751492	C	T	0.26	0.74	0.72	0.28	178.91	8.38E-41	4.94E-39	Alt	EUR
1	66089782	rs6700896	C	T	0.39	0.61	0.23	0.77	26.09	3.26E-07	2.37E-06	Alt	NAT
1	70887099	rs672203	A	G	0.91	0.09	0.74	0.26	25.42	4.62E-07	3.28E-06	Ref	NAT
1	71514969	rs7551789	A	T	0.93	0.07	0.77	0.23	25.63	4.13E-07	2.95E-06	Ref	EUR
1	72765116	rs2568958	G	A	0.28	0.72	0.13	0.87	31.43	2.07E-08	1.75E-07	Alt	NAT
1	76839536	rs12144344	C	T	0.83	0.17	0.71	0.29	11.47	7.08E-04	3.03E-03	Ref	AFR
1	89849574	rs928655	G	A	0.44	0.56	0.61	0.39	19.31	1.11E-05	6.46E-05	Alt	EUR
1	94674726	rs2274788	T	C	0.80	0.20	0.67	0.33	12.89	3.31E-04	1.51E-03	Ref	AFR
1	95053353	rs12029080	T	G	0.82	0.18	0.69	0.31	13.27	2.70E-04	1.25E-03	Ref	AFR
1	96943994	rs1973993	T	C	0.43	0.57	0.19	0.81	64.71	8.67E-16	1.51E-14	Alt	NAT
1	97037083	rs10783050	T	C	0.75	0.25	0.55	0.45	27.44	1.62E-07	1.22E-06	Ref	AFR
1	98348885	rs1801265	G	A	0.16	0.84	0.10	0.90	7.91	4.92E-03	1.77E-02	Alt	NAT
1	99782957	rs12743824	C	A	0.26	0.74	0.37	0.63	9.10	2.55E-03	9.86E-03	Alt	NAT
1	109818306	rs629301	G	T	0.22	0.78	0.07	0.93	60.61	6.95E-15	1.12E-13	Alt	NAT
1	109818530	rs646776	C	T	0.22	0.78	0.07	0.93	60.61	6.95E-15	1.12E-13	Alt	NAT
1	109822166	rs599839	G	A	0.24	0.76	0.09	0.91	49.43	2.06E-12	2.68E-11	Alt	NAT
1	115837709	rs2239622	A	G	0.19	0.81	0.40	0.60	30.18	3.94E-08	3.21E-07	Alt	AFR
1	149892872	rs11205277	A	G	0.74	0.26	0.63	0.37	9.10	2.55E-03	9.86E-03	Ref	AFR
1	153769400	rs9426935	C	T	0.79	0.21	0.86	0.14	5.87	1.54E-02	4.98E-02	Ref	NAT
1	154418879	rs4537545	C	T	0.28	0.72	0.60	0.40	74.52	6.01E-18	1.22E-16	Alt	AFR
1	155279482	rs2297480	T	G	0.69	0.31	0.34	0.66	91.10	1.37E-21	3.53E-20	Ref	EUR

Table 5 (continued).

1	155868625	rs2282301	G	A	0.73	0.27	0.37	0.63	97.59	5.14E-23	1.43E-21	Ref	EUR
1	156873727	rs12407843	G	A	0.87	0.13	0.67	0.33	30.78	2.89E-08	2.39E-07	Ref	EUR
1	156877797	rs77235035	C	A	0.83	0.17	0.66	0.34	22.01	2.71E-06	1.73E-05	Ref	EUR
1	156879580	rs3737224	C	T	0.92	0.08	0.71	0.29	36.72	1.36E-09	1.32E-08	Ref	EUR
1	156881959	rs41273215	C	T	0.92	0.08	0.70	0.30	36.72	1.36E-09	1.32E-08	Ref	EUR
1	156883215	rs822442	C	A	0.92	0.08	0.72	0.28	33.56	6.91E-09	6.15E-08	Ref	EUR
1	159698549	rs7553007	G	A	0.61	0.39	0.50	0.50	8.49	3.56E-03	1.32E-02	Ref	AFR
1	161479745	rs1801274	A	G	0.53	0.47	0.63	0.37	7.29	6.94E-03	2.41E-02	Ref	NAT
1	162033890	rs12143842	C	T	0.88	0.12	0.67	0.33	32.62	1.12E-08	9.75E-08	Ref	AFR
1	162085685	rs10494366	G	T	0.39	0.61	0.63	0.37	42.68	6.45E-11	7.26E-10	Alt	EUR
1	162210610	rs4657178	C	T	0.66	0.34	0.52	0.48	13.58	2.28E-04	1.07E-03	Ref	EUR
1	165448157	rs10918196	T	C	0.49	0.51	0.25	0.75	52.33	4.69E-13	6.51E-12	Alt	NAT
1	169580290	rs2235302	C	T	0.89	0.11	0.76	0.24	15.56	8.01E-05	4.04E-04	Ref	NAT
1	171076966	rs2266782	G	A	0.70	0.30	0.79	0.21	7.93	4.86E-03	1.75E-02	Ref	NAT
1	171080080	rs1736557	G	A	0.89	0.11	0.77	0.23	15.82	6.95E-05	3.54E-04	Ref	AFR
1	171083242	rs2266780	A	G	0.94	0.06	0.84	0.16	12.72	3.61E-04	1.63E-03	Ref	AFR
1	171091875	rs1795240	A	G	0.15	0.85	0.39	0.61	39.63	3.07E-10	3.20E-09	Alt	AFR
1	171254890	rs7877	C	T	0.54	0.46	0.76	0.24	46.41	9.58E-12	1.16E-10	Ref	NAT
1	172189889	rs678962	T	G	0.78	0.22	0.68	0.32	7.84	5.10E-03	1.83E-02	Ref	EUR
1	186947224	rs10157410	G	C	0.76	0.24	0.86	0.14	12.42	4.25E-04	1.89E-03	Ref	NAT
1	196679455	rs10737680	A	C	0.42	0.58	0.62	0.38	30.70	3.02E-08	2.49E-07	Alt	AFR
1	201345487	rs12564445	G	A	0.86	0.14	0.67	0.33	27.52	1.55E-07	1.18E-06	Ref	EUR
1	203135452	rs16851030	C	T	0.86	0.14	0.70	0.30	23.30	1.39E-06	9.18E-06	Ref	EUR
1	206946407	rs1800872	T	G	0.37	0.63	0.65	0.35	59.81	1.05E-14	1.67E-13	Alt	EUR
1	206946634	rs1800871	A	G	0.37	0.63	0.65	0.35	59.81	1.05E-14	1.67E-13	Alt	EUR

Table 5 (continued).

1	206946897	rs1800896	T	C	0.75	0.25	0.89	0.11	35.79	2.20E-09	2.07E-08	Ref	NAT
1	210536025	rs7527939	C	T	0.71	0.29	0.87	0.13	42.60	6.73E-11	7.55E-10	Ref	NAT
1	218931905	rs12037343	G	T	0.91	0.09	0.72	0.28	30.11	4.08E-08	3.31E-07	Ref	AFR
1	230294916	rs2144300	C	T	0.45	0.55	0.74	0.26	73.62	9.45E-18	1.90E-16	Alt	EUR
1	230838331	rs7079	G	T	0.91	0.09	0.82	0.18	9.11	2.55E-03	9.85E-03	Ref	AFR
1	232756026	rs10495332	T	C	0.91	0.09	0.74	0.26	24.04	9.43E-07	6.38E-06	Ref	AFR
1	233719984	rs11800854	G	A	0.95	0.05	0.87	0.13	9.18	2.44E-03	9.47E-03	Ref	EUR
1	237028564	rs3768142	G	T	0.28	0.72	0.41	0.59	10.76	1.04E-03	4.32E-03	Alt	AFR
1	247675559	rs7550918	C	T	0.23	0.77	0.05	0.95	105.59	9.07E-25	2.79E-23	Alt	NAT
2	10903412	rs1198872	C	T	0.72	0.28	0.54	0.46	22.77	1.83E-06	1.20E-05	Ref	EUR
2	21225281	rs1042034	C	T	0.38	0.62	0.59	0.41	33.40	7.52E-09	6.67E-08	Alt	AFR
2	21231524	rs676210	G	A	0.62	0.38	0.41	0.59	33.40	7.52E-09	6.67E-08	Ref	AFR
2	25131316	rs6545814	A	G	0.80	0.20	0.57	0.43	38.29	6.10E-10	6.16E-09	Ref	NAT
2	27730940	rs1260326	T	C	0.32	0.68	0.55	0.45	36.19	1.79E-09	1.71E-08	Alt	AFR
2	27741237	rs780094	T	C	0.33	0.67	0.56	0.44	34.36	4.58E-09	4.17E-08	Alt	AFR
2	27742603	rs780093	T	C	0.33	0.67	0.56	0.44	34.36	4.58E-09	4.17E-08	Alt	AFR
2	31247514	rs9679162	G	T	0.32	0.68	0.48	0.52	17.17	3.41E-05	1.85E-04	Alt	AFR
2	31249427	rs12613732	T	G	0.38	0.62	0.58	0.42	26.24	3.02E-07	2.21E-06	Alt	NAT
2	41273631	rs17027130	T	C	0.88	0.12	0.73	0.27	18.63	1.59E-05	9.04E-05	Ref	AFR
2	56120853	rs1430193	A	T	0.62	0.38	0.18	0.82	221.92	3.45E-50	2.62E-48	Ref	EUR
2	56195696	rs62164511	A	G	0.34	0.66	0.84	0.16	301.69	1.41E-67	1.59E-65	Alt	EUR
2	60459806	rs359268	T	C	0.46	0.54	0.60	0.40	14.12	1.72E-04	8.23E-04	Alt	EUR
2	64503895	rs2698530	A	C	0.65	0.35	0.46	0.54	25.85	3.69E-07	2.66E-06	Ref	AFR
2	69723710	rs7577851	C	T	0.61	0.39	0.70	0.30	7.17	7.41E-03	2.56E-02	Ref	EUR
2	105897740	rs1020064	T	G	0.04	0.96	0.22	0.78	32.57	1.15E-08	9.99E-08	Alt	AFR

Table 5 (continued).

2	111907691	rs724710	T	C	0.53	0.47	0.19	0.81	122.17	2.12E-28	7.77E-27	Ref	EUR
2	113594867	rs16944	A	G	0.70	0.30	0.43	0.57	50.80	1.02E-12	1.38E-11	Ref	AFR
2	113598107	rs4848306	G	A	0.75	0.25	0.50	0.50	43.51	4.21E-11	4.82E-10	Ref	AFR
2	128018063	rs3738948	A	G	0.91	0.09	0.69	0.31	39.59	3.14E-10	3.26E-09	Ref	AFR
2	130468366	rs10928927	C	T	0.34	0.66	0.21	0.79	17.06	3.63E-05	1.95E-04	Alt	EUR
2	134266001	rs16826005	A	G	0.95	0.05	0.76	0.24	32.91	9.63E-09	8.45E-08	Ref	EUR
2	152981335	rs16830728	G	T	0.95	0.05	0.47	0.53	154.91	1.46E-35	7.23E-34	Ref	EUR
2	159899489	rs7582141	G	T	0.80	0.20	0.89	0.11	15.91	6.66E-05	3.40E-04	Ref	EUR
2	159899913	rs6432512	C	T	0.80	0.20	0.89	0.11	15.91	6.66E-05	3.40E-04	Ref	EUR
2	159936391	rs264588	C	A	0.80	0.20	0.91	0.09	26.40	2.78E-07	2.04E-06	Ref	EUR
2	159950865	rs264631	C	G	0.79	0.21	0.91	0.09	29.25	6.37E-08	5.06E-07	Ref	EUR
2	162910223	rs6741949	G	C	0.42	0.58	0.88	0.12	353.66	6.77E-79	9.33E-77	Alt	EUR
2	162997960	rs13429709	T	C	0.16	0.84	0.40	0.60	39.22	3.79E-10	3.92E-09	Alt	NAT
2	165513091	rs10195252	T	C	0.90	0.10	0.76	0.24	17.29	3.20E-05	1.74E-04	Ref	NAT
2	166168503	rs2304016	A	G	0.97	0.03	0.88	0.12	12.75	3.56E-04	1.61E-03	Ref	AFR
2	169041386	rs6749447	T	G	0.46	0.54	0.36	0.64	7.39	6.56E-03	2.29E-02	Alt	NAT
2	170010985	rs2075252	T	C	0.24	0.76	0.49	0.51	41.53	1.16E-10	1.27E-09	Alt	AFR
2	174504924	rs13028485	G	A	0.95	0.05	0.80	0.20	24.32	8.15E-07	5.56E-06	Ref	EUR
2	176289319	rs2461751	G	A	0.79	0.21	0.56	0.44	40.00	2.54E-10	2.67E-09	Ref	EUR
2	199632565	rs12617311	G	A	0.71	0.29	0.61	0.39	6.34	1.18E-02	3.92E-02	Ref	AFR
2	200638509	rs12615435	T	G	0.93	0.07	0.81	0.19	15.40	8.71E-05	4.37E-04	Ref	AFR
2	211060050	rs2286963	T	G	0.92	0.08	0.79	0.21	15.87	6.80E-05	3.46E-04	Ref	AFR
2	213824045	rs12619285	A	G	0.26	0.74	0.44	0.56	21.54	3.47E-06	2.18E-05	Alt	NAT
2	216205167	rs16853826	G	A	0.79	0.21	0.59	0.41	28.07	1.17E-07	8.97E-07	Ref	EUR
2	233513175	rs2140773	C	A	0.56	0.44	0.67	0.33	8.55	3.45E-03	1.28E-02	Ref	NAT

Table 5 (continued).

2	234526871	rs1042597	C	G	0.75	0.25	0.54	0.46	29.02	7.16E-08	5.65E-07	Ref	AFR
2	234529643	rs6431558	C	T	0.71	0.29	0.43	0.57	53.03	3.28E-13	4.63E-12	Ref	EUR
2	234579892	rs3806598	A	C	0.99	0.01	0.80	0.20	40.04	2.49E-10	2.63E-09	Ref	EUR
2	234681544	rs1042640	G	C	0.21	0.79	0.14	0.86	5.87	1.54E-02	4.98E-02	Alt	NAT
3	2624938	rs2619566	G	A	0.54	0.46	0.36	0.64	24.57	7.16E-07	4.92E-06	Ref	AFR
3	12267648	rs7616006	A	G	0.89	0.11	0.61	0.39	59.15	1.46E-14	2.30E-13	Ref	NAT
3	34865597	rs559356	T	C	0.75	0.25	0.52	0.48	34.05	5.37E-09	4.85E-08	Ref	AFR
3	38442490	rs2070488	G	A	0.80	0.20	0.71	0.29	6.45	1.11E-02	3.71E-02	Ref	AFR
3	38594973	rs7638909	T	G	0.60	0.40	0.49	0.51	8.50	3.55E-03	1.32E-02	Ref	EUR
3	38624253	rs3922844	T	C	0.34	0.66	0.19	0.81	26.02	3.37E-07	2.44E-06	Alt	NAT
3	45731451	rs2742417	C	T	0.62	0.38	0.72	0.28	8.39	3.77E-03	1.39E-02	Ref	NAT
3	45731784	rs2251954	T	C	0.62	0.38	0.72	0.28	9.53	2.02E-03	7.99E-03	Ref	NAT
3	45732515	rs2742421	T	G	0.62	0.38	0.72	0.28	9.53	2.02E-03	7.99E-03	Ref	NAT
3	45733430	rs2742423	A	G	0.62	0.38	0.72	0.28	9.53	2.02E-03	7.99E-03	Ref	NAT
3	45734818	rs1969624	T	C	0.62	0.38	0.72	0.28	9.53	2.02E-03	7.99E-03	Ref	NAT
3	45740863	rs2742435	G	A	0.64	0.36	0.76	0.24	14.17	1.67E-04	7.99E-04	Ref	AFR
3	45749722	rs2245705	T	C	0.62	0.38	0.73	0.27	10.76	1.04E-03	4.32E-03	Ref	NAT
3	45756722	rs2742390	G	A	0.62	0.38	0.73	0.27	10.76	1.04E-03	4.32E-03	Ref	NAT
3	53680124	rs1401492	C	T	0.98	0.02	0.91	0.09	9.94	1.62E-03	6.52E-03	Ref	EUR
3	56849749	rs1354034	T	C	0.66	0.34	0.49	0.51	19.79	8.64E-06	5.10E-05	Ref	AFR
3	113046640	rs13064411	A	G	0.96	0.04	0.91	0.09	6.90	8.62E-03	2.95E-02	Ref	AFR
3	113879562	rs167770	G	A	0.53	0.47	0.31	0.69	37.53	9.00E-10	8.87E-09	Ref	AFR
3	113885395	rs324023	T	C	0.53	0.47	0.31	0.69	37.53	9.00E-10	8.87E-09	Ref	AFR
3	113890815	rs6280	C	T	0.54	0.46	0.33	0.67	34.51	4.24E-09	3.88E-08	Ref	AFR
3	113891042	rs324026	C	T	0.54	0.46	0.33	0.67	34.51	4.24E-09	3.88E-08	Ref	AFR

Table 5 (continued).

3	117574822	rs6438424	A	C	0.65	0.35	0.50	0.50	15.91	6.65E-05	3.40E-04	Ref	AFR
3	119499507	rs1523130	T	C	0.42	0.58	0.31	0.69	9.90	1.66E-03	6.64E-03	Alt	NAT
3	119525497	rs7643645	A	G	0.36	0.64	0.57	0.43	32.98	9.30E-09	8.16E-08	Alt	NAT
3	119537291	rs3814058	T	C	0.82	0.18	0.59	0.41	36.94	1.22E-09	1.19E-08	Ref	EUR
3	119631814	rs6438552	A	G	0.71	0.29	0.47	0.53	39.69	2.98E-10	3.11E-09	Ref	EUR
3	119813282	rs334558	A	G	0.73	0.27	0.44	0.56	57.29	3.77E-14	5.73E-13	Ref	EUR
3	141102833	rs6763931	G	A	0.51	0.49	0.64	0.36	12.37	4.35E-04	1.94E-03	Ref	NAT
3	151056598	rs6809699	A	C	0.07	0.93	0.14	0.86	6.08	1.36E-02	4.46E-02	Alt	AFR
3	151090996	rs9859552	G	T	0.93	0.07	0.97	0.03	10.10	1.49E-03	6.02E-03	Ref	AFR
3	152672779	rs6785504	G	T	0.23	0.77	0.53	0.47	63.94	1.28E-15	2.19E-14	Alt	AFR
3	160820524	rs4557202	G	C	0.44	0.56	0.23	0.77	40.76	1.72E-10	1.84E-09	Alt	NAT
3	167837748	rs4345115	T	C	0.39	0.61	0.54	0.46	16.01	6.29E-05	3.23E-04	Alt	NAT
3	184010048	rs3914188	G	C	0.15	0.85	0.38	0.62	38.11	6.67E-10	6.71E-09	Alt	EUR
3	184090266	rs6141	C	T	0.26	0.74	0.51	0.49	39.56	3.18E-10	3.31E-09	Alt	AFR
3	185629568	rs2002675	A	G	0.91	0.09	0.82	0.18	7.93	4.85E-03	1.75E-02	Ref	NAT
3	187456709	rs3733017	T	G	0.94	0.06	0.84	0.16	12.36	4.39E-04	1.95E-03	Ref	EUR
4	2906707	rs4961	G	T	0.81	0.19	0.54	0.46	52.24	4.92E-13	6.81E-12	Ref	AFR
4	3006043	rs1024323	C	T	0.65	0.35	0.76	0.24	10.41	1.25E-03	5.13E-03	Ref	NAT
4	4252956	rs2980098	A	G	0.33	0.67	0.63	0.37	65.59	5.54E-16	9.80E-15	Alt	AFR
4	8503359	rs1949733	A	G	0.51	0.49	0.37	0.63	14.53	1.38E-04	6.72E-04	Ref	NAT
4	9926967	rs13129697	T	G	0.26	0.74	0.57	0.43	67.68	1.93E-16	3.54E-15	Alt	AFR
4	16893893	rs1483012	G	A	0.61	0.39	0.78	0.22	31.09	2.46E-08	2.06E-07	Ref	NAT
4	16908004	rs6819013	A	G	0.68	0.32	0.89	0.11	72.59	1.60E-17	3.16E-16	Ref	NAT
4	18017730	rs6830062	T	C	0.93	0.07	0.86	0.14	7.93	4.87E-03	1.75E-02	Ref	EUR
4	38139024	rs9852	C	T	0.97	0.03	0.69	0.31	63.16	1.90E-15	3.21E-14	Ref	AFR

Table 5 (continued).

4	69681936	rs61750900	G	T	0.96	0.04	0.99	0.01	8.10	4.44E-03	1.60E-02	Ref	NAT
4	69962078	rs7668258	T	C	0.14	0.86	0.33	0.67	28.74	8.27E-08	6.44E-07	Alt	AFR
4	71859193	rs80143932	C	G	0.94	0.06	0.86	0.14	9.51	2.04E-03	8.06E-03	Ref	AFR
4	71859352	rs2306744	C	T	0.94	0.06	0.86	0.14	9.51	2.04E-03	8.06E-03	Ref	AFR
4	74606024	rs4073	A	T	0.27	0.73	0.44	0.56	18.76	1.48E-05	8.48E-05	Alt	NAT
4	74864550	rs352046	G	C	0.09	0.91	0.04	0.96	12.07	5.13E-04	2.25E-03	Alt	NAT
4	78605369	rs958617	A	G	0.46	0.54	0.36	0.64	7.39	6.56E-03	2.29E-02	Alt	EUR
4	84192168	rs4693075	G	C	0.54	0.46	0.16	0.84	169.70	8.60E-39	4.73E-37	Ref	AFR
4	88213808	rs6834314	A	G	0.94	0.06	0.69	0.31	50.69	1.08E-12	1.45E-11	Ref	AFR
4	89030841	rs12505410	T	G	0.46	0.54	0.70	0.30	47.09	6.79E-12	8.39E-11	Alt	EUR
4	89052323	rs2231142	G	T	0.85	0.15	0.74	0.26	11.07	8.78E-04	3.70E-03	Ref	AFR
4	89096061	rs7699188	G	A	0.95	0.05	0.88	0.12	8.16	4.28E-03	1.56E-02	Ref	NAT
4	100163873	rs2201728	G	A	0.31	0.69	0.79	0.21	231.21	3.24E-52	2.64E-50	Alt	EUR
4	100239319	rs1229984	T	C	0.01	0.99	0.58	0.42	222.04	3.25E-50	2.46E-48	Alt	AFR
4	100495488	rs1800591	G	T	0.91	0.09	0.83	0.17	7.03	8.03E-03	2.76E-02	Ref	NAT
4	114861300	rs4460079	C	T	0.35	0.65	0.25	0.75	10.25	1.37E-03	5.59E-03	Alt	NAT
4	142709723	rs17007695	T	C	0.44	0.56	0.64	0.36	31.32	2.19E-08	1.84E-07	Alt	NAT
4	146794621	rs4547811	T	C	0.38	0.62	0.63	0.37	46.94	7.31E-12	9.00E-11	Alt	NAT
4	146875551	rs723794	G	T	0.19	0.81	0.47	0.53	54.33	1.69E-13	2.44E-12	Alt	EUR
4	154609523	rs1816702	T	C	0.21	0.79	0.03	0.97	198.02	5.64E-45	3.79E-43	Alt	NAT
4	155514879	rs13109457	G	A	0.82	0.18	0.54	0.46	54.48	1.57E-13	2.27E-12	Ref	EUR
4	159630817	rs8396	T	C	0.95	0.05	0.82	0.18	20.87	4.91E-06	3.01E-05	Ref	NAT
4	175071602	rs12507634	A	G	0.39	0.61	0.78	0.22	155.08	1.35E-35	6.66E-34	Alt	EUR
4	182197947	rs1454694	T	C	0.95	0.05	0.80	0.20	26.23	3.03E-07	2.22E-06	Ref	AFR
5	610093	rs924607	C	T	0.71	0.29	0.60	0.40	7.94	4.83E-03	1.74E-02	Ref	AFR

Table 5 (continued).

5	1322087	rs401681	C	T	0.58	0.42	0.69	0.31	11.08	8.72E-04	3.68E-03	Ref	NAT
5	8652870	rs200113	T	C	0.97	0.03	0.78	0.22	35.37	2.72E-09	2.54E-08	Ref	AFR
5	17003085	rs6870564	A	G	0.05	0.95	0.14	0.86	12.42	4.25E-04	1.89E-03	Alt	EUR
5	37818139	rs2973049	T	C	0.33	0.67	0.44	0.56	7.75	5.36E-03	1.91E-02	Alt	AFR
5	37828844	rs2216711	G	A	0.11	0.89	0.29	0.71	27.23	1.81E-07	1.35E-06	Alt	AFR
5	57214817	rs10041935	A	C	0.82	0.18	0.67	0.33	16.64	4.51E-05	2.39E-04	Ref	AFR
5	59736773	rs702553	A	T	0.71	0.29	0.45	0.55	46.07	1.14E-11	1.38E-10	Ref	EUR
5	63250851	rs1364043	T	G	0.68	0.32	0.51	0.49	19.80	8.60E-06	5.08E-05	Ref	AFR
5	73276903	rs6894385	A	C	0.86	0.14	0.74	0.26	12.27	4.61E-04	2.05E-03	Ref	AFR
5	75514986	rs11960832	C	T	0.51	0.49	0.66	0.34	17.69	2.60E-05	1.43E-04	Ref	NAT
5	76224171	rs2460504	G	C	0.84	0.16	0.68	0.32	19.78	8.67E-06	5.12E-05	Ref	AFR
5	76781471	rs163030	A	C	0.13	0.87	0.42	0.58	60.23	8.43E-15	1.35E-13	Alt	AFR
5	79950781	rs1650697	A	G	0.28	0.72	0.37	0.63	6.46	1.11E-02	3.70E-02	Alt	AFR
5	79951496	rs442767	G	T	0.39	0.61	0.52	0.48	11.41	7.29E-04	3.12E-03	Alt	NAT
5	88183651	rs17560407	A	G	0.94	0.06	0.76	0.24	29.42	5.82E-08	4.64E-07	Ref	AFR
5	122685128	rs2115172	A	G	0.61	0.39	0.51	0.49	7.62	5.76E-03	2.04E-02	Ref	EUR
5	131819921	rs2070729	C	A	0.24	0.76	0.40	0.60	19.22	1.17E-05	6.79E-05	Alt	AFR
5	132009154	rs2243250	C	T	0.52	0.48	0.37	0.63	15.76	7.18E-05	3.65E-04	Ref	EUR
5	133849177	rs13187289	C	G	0.85	0.15	0.94	0.06	21.87	2.92E-06	1.85E-05	Ref	NAT
5	137707315	rs757647	G	A	0.66	0.34	0.48	0.52	22.64	1.95E-06	1.27E-05	Ref	EUR
5	159820931	rs10515808	C	A	0.94	0.06	0.98	0.02	9.22	2.40E-03	9.32E-03	Ref	NAT
5	161324898	rs2290732	A	G	0.26	0.74	0.49	0.51	35.81	2.18E-09	2.05E-08	Alt	EUR
5	167500460	rs13358864	T	A	0.95	0.05	0.80	0.20	24.85	6.19E-07	4.30E-06	Ref	AFR
5	174876002	rs2168631	G	A	0.81	0.19	0.64	0.36	18.78	1.47E-05	8.42E-05	Ref	EUR
5	176836532	rs1801020	A	G	0.51	0.49	0.62	0.38	8.07	4.50E-03	1.63E-02	Ref	NAT

Table 5 (continued).

6	2235633	rs9378688	G	A	0.94	0.06	0.84	0.16	11.27	7.87E-04	3.34E-03	Ref	AFR
6	7102084	rs675209	T	C	0.62	0.38	0.79	0.21	27.63	1.47E-07	1.12E-06	Ref	NAT
6	7801112	rs6923462	T	C	0.84	0.16	0.97	0.03	109.01	1.62E-25	5.16E-24	Ref	NAT
6	18139802	rs12201199	A	T	0.90	0.10	0.97	0.03	43.27	4.77E-11	5.43E-10	Ref	NAT
6	20685486	rs9356744	T	C	0.71	0.29	0.62	0.38	6.38	1.16E-02	3.84E-02	Ref	EUR
6	24491475	rs1883415	A	C	0.92	0.08	0.82	0.18	10.36	1.29E-03	5.25E-03	Ref	NAT
6	24503590	rs2760118	C	T	0.91	0.09	0.82	0.18	7.93	4.85E-03	1.75E-02	Ref	NAT
6	31241109	rs13191343	C	T	0.94	0.06	0.87	0.13	6.32	1.20E-02	3.96E-02	Ref	AFR
6	31253444	rs9461684	C	T	0.78	0.22	0.91	0.09	35.39	2.70E-09	2.52E-08	Ref	NAT
6	31274380	rs9264942	T	C	0.78	0.22	0.57	0.43	29.41	5.86E-08	4.67E-07	Ref	AFR
6	31322559	rs2523608	G	A	0.39	0.61	0.49	0.51	6.80	9.11E-03	3.09E-02	Alt	EUR
6	31431691	rs2255221	G	T	0.93	0.07	0.97	0.03	6.22	1.26E-02	4.16E-02	Ref	NAT
6	32336187	rs3129934	T	C	0.04	0.96	0.10	0.90	6.54	1.06E-02	3.55E-02	Alt	NAT
6	32632832	rs9274407	A	T	0.13	0.87	0.26	0.74	15.99	6.38E-05	3.27E-04	Alt	EUR
6	32678999	rs9275572	A	G	0.22	0.78	0.32	0.68	7.77	5.32E-03	1.89E-02	Alt	NAT
6	32972404	rs3128935	T	C	0.93	0.07	0.85	0.15	7.93	4.87E-03	1.75E-02	Ref	EUR
6	32975014	rs399604	T	C	0.64	0.36	0.48	0.52	15.93	6.59E-05	3.37E-04	Ref	AFR
6	33048661	rs1042151	A	G	0.72	0.28	0.89	0.11	46.46	9.37E-12	1.14E-10	Ref	NAT
6	33055538	rs9277554	C	T	0.65	0.35	0.49	0.51	15.91	6.65E-05	3.40E-04	Ref	EUR
6	33465482	rs9394152	C	T	0.34	0.66	0.17	0.83	34.96	3.36E-09	3.11E-08	Alt	NAT
6	35369806	rs1883322	C	T	0.15	0.85	0.29	0.71	16.32	5.35E-05	2.79E-04	Alt	NAT
6	35395010	rs3734254	C	T	0.15	0.85	0.27	0.73	13.14	2.89E-04	1.33E-03	Alt	NAT
6	35402785	rs4713858	A	G	0.09	0.91	0.20	0.80	11.91	5.58E-04	2.43E-03	Alt	AFR
6	39307032	rs9471075	A	C	0.94	0.06	0.98	0.02	16.63	4.55E-05	2.40E-04	Ref	NAT
6	39325078	rs20455	A	G	0.79	0.21	0.53	0.47	47.81	4.69E-12	5.89E-11	Ref	EUR

Table 5 (continued).

6	45095163	rs9395066	A	C	0.58	0.42	0.35	0.65	41.53	1.16E-10	1.27E-09	Ref	AFR
6	53924697	rs9296736	T	C	0.69	0.31	0.47	0.53	34.15	5.11E-09	4.64E-08	Ref	AFR
6	55639028	rs41271330	G	A	0.92	0.08	0.86	0.14	5.87	1.54E-02	4.98E-02	Ref	AFR
6	79556166	rs16890334	T	C	0.99	0.01	0.94	0.06	8.61	3.35E-03	1.25E-02	Ref	AFR
6	81302805	rs10943724	C	A	0.44	0.56	0.54	0.46	6.85	8.88E-03	3.03E-02	Alt	NAT
6	107660646	rs2430457	G	A	0.02	0.98	0.01	0.99	9.05	2.62E-03	9.98E-03	Alt	NAT
6	110777962	rs6907567	A	G	0.80	0.20	0.55	0.45	41.98	9.24E-11	1.02E-09	Ref	EUR
6	110778128	rs714368	T	C	0.80	0.20	0.56	0.44	40.11	2.40E-10	2.54E-09	Ref	EUR
6	111922503	rs76228616	G	C	0.91	0.09	0.99	0.01	197.16	8.70E-45	5.72E-43	Ref	NAT
6	118574061	rs281868	G	A	0.85	0.15	0.68	0.32	21.28	3.97E-06	2.47E-05	Ref	NAT
6	118680374	rs11970286	C	T	0.91	0.09	0.70	0.30	36.30	1.69E-09	1.62E-08	Ref	AFR
6	121748542	rs11154022	A	G	0.26	0.74	0.49	0.51	34.00	5.52E-09	4.98E-08	Alt	AFR
6	122146034	rs9398652	C	A	0.94	0.06	0.65	0.35	62.33	2.91E-15	4.87E-14	Ref	EUR
6	126964510	rs4273712	A	G	0.44	0.56	0.55	0.45	8.59	3.38E-03	1.26E-02	Alt	NAT
6	135419018	rs9399137	T	C	0.90	0.10	0.72	0.28	26.47	2.68E-07	1.97E-06	Ref	AFR
6	135432552	rs9494145	T	C	0.87	0.13	0.76	0.24	11.60	6.58E-04	2.83E-03	Ref	AFR
6	137102365	rs9376230	C	A	0.31	0.69	0.42	0.58	8.70	3.19E-03	1.19E-02	Alt	AFR
6	137673302	rs6928289	G	A	0.52	0.48	0.32	0.68	31.07	2.49E-08	2.08E-07	Ref	EUR
6	142512136	rs225675	A	G	0.87	0.13	0.79	0.21	6.91	8.59E-03	2.94E-02	Ref	AFR
6	147680359	rs9390459	A	G	0.42	0.58	0.67	0.33	46.56	8.89E-12	1.09E-10	Alt	AFR
6	154360797	rs1799971	A	G	0.80	0.20	0.69	0.31	9.90	1.66E-03	6.64E-03	Ref	AFR
6	154487421	rs2281617	C	T	0.93	0.07	0.60	0.40	76.86	1.83E-18	3.88E-17	Ref	EUR
6	160551204	rs683369	G	C	0.04	0.96	0.14	0.86	14.02	1.81E-04	8.63E-04	Alt	AFR
6	160557643	rs2282143	C	T	0.99	0.01	0.89	0.11	15.91	6.66E-05	3.40E-04	Ref	EUR
6	160670282	rs316019	A	C	0.05	0.95	0.13	0.87	8.82	2.97E-03	1.12E-02	Alt	EUR

Table 5 (continued).

6	166579270	rs2305089	C	T	0.41	0.59	0.64	0.36	38.89	4.49E-10	4.61E-09	Alt	EUR
7	1976457	rs1801368	C	T	0.45	0.55	0.55	0.45	7.71	5.49E-03	1.95E-02	Alt	NAT
7	7313485	rs6953213	G	A	0.92	0.08	0.83	0.17	10.64	1.10E-03	4.58E-03	Ref	EUR
7	28004198	rs4722750	C	T	0.97	0.03	0.84	0.16	21.31	3.91E-06	2.43E-05	Ref	EUR
7	30699972	rs2270007	G	C	0.14	0.86	0.40	0.60	49.63	1.85E-12	2.43E-11	Alt	AFR
7	30726777	rs7793837	A	T	0.70	0.30	0.80	0.20	9.21	2.41E-03	9.35E-03	Ref	NAT
7	32444435	rs215738	G	A	0.10	0.90	0.02	0.98	43.27	4.77E-11	5.43E-10	Alt	NAT
7	33060946	rs3750117	A	G	0.26	0.74	0.49	0.51	37.65	8.45E-10	8.38E-09	Alt	AFR
7	41470093	rs1079866	C	G	0.89	0.11	0.80	0.20	9.41	2.16E-03	8.42E-03	Ref	AFR
7	73020337	rs3812316	C	G	0.96	0.04	0.88	0.12	11.11	8.60E-04	3.63E-03	Ref	AFR
7	74126034	rs117026326	C	T	0.99	0.01	0.91	0.09	15.52	8.15E-05	4.10E-04	Ref	AFR
7	80236014	rs13236689	T	G	0.56	0.44	0.44	0.56	11.58	6.66E-04	2.86E-03	Ref	EUR
7	80532112	rs7779029	T	C	0.98	0.02	0.89	0.11	16.38	5.19E-05	2.72E-04	Ref	EUR
7	87133366	rs3842	T	C	0.84	0.16	0.73	0.27	10.76	1.04E-03	4.32E-03	Ref	AFR
7	87157051	rs7787082	G	A	0.73	0.27	0.55	0.45	21.42	3.70E-06	2.31E-05	Ref	EUR
7	87164986	rs10248420	A	G	0.73	0.27	0.56	0.44	20.07	7.48E-06	4.47E-05	Ref	EUR
7	87171152	rs4148737	T	C	0.43	0.57	0.69	0.31	56.10	6.88E-14	1.02E-12	Alt	EUR
7	87173667	rs1922242	A	T	0.43	0.57	0.69	0.31	56.10	6.88E-14	1.02E-12	Alt	EUR
7	87179601	rs1128503	A	G	0.33	0.67	0.63	0.37	65.59	5.54E-16	9.80E-15	Alt	AFR
7	87180198	rs10276036	C	T	0.33	0.67	0.63	0.37	65.59	5.54E-16	9.80E-15	Alt	AFR
7	87230193	rs3213619	A	G	0.99	0.01	0.94	0.06	7.87	5.02E-03	1.80E-02	Ref	EUR
7	92264410	rs2282978	T	C	0.62	0.38	0.85	0.15	65.57	5.62E-16	9.94E-15	Ref	NAT
7	94925820	rs854548	A	G	0.52	0.48	0.23	0.77	83.19	7.47E-20	1.73E-18	Ref	AFR
7	94930391	rs854555	A	C	0.62	0.38	0.52	0.48	7.63	5.73E-03	2.03E-02	Ref	AFR
7	99270539	rs776746	C	T	0.88	0.12	0.72	0.28	19.88	8.25E-06	4.89E-05	Ref	EUR

Table 5 (continued).

7	99361466	rs2242480	C	T	0.42	0.58	0.76	0.24	109.98	9.91E-26	3.21E-24	Alt	AFR
7	99365083	rs4646437	G	A	0.95	0.05	0.85	0.15	13.12	2.92E-04	1.34E-03	Ref	EUR
7	101809851	rs365836	A	G	0.73	0.27	0.87	0.13	26.60	2.50E-07	1.85E-06	Ref	NAT
7	123411223	rs4731120	A	C	0.97	0.03	0.90	0.10	8.35	3.86E-03	1.42E-02	Ref	AFR
7	127164958	rs6467136	A	G	0.56	0.44	0.27	0.73	75.56	3.55E-18	7.35E-17	Ref	EUR
7	150696111	rs1799983	T	G	0.09	0.91	0.18	0.82	9.11	2.55E-03	9.85E-03	Alt	AFR
7	152878333	rs4285401	A	G	0.80	0.20	0.66	0.34	13.96	1.87E-04	8.87E-04	Ref	AFR
7	154509324	rs12666280	T	C	0.34	0.66	0.57	0.43	38.55	5.34E-10	5.43E-09	Alt	EUR
8	1244224	rs17669535	C	G	0.98	0.02	0.86	0.14	21.40	3.74E-06	2.33E-05	Ref	AFR
8	2740502	rs641525	T	G	0.82	0.18	0.70	0.30	13.41	2.50E-04	1.17E-03	Ref	EUR
8	4078353	rs2407314	G	C	0.40	0.60	0.57	0.43	21.68	3.22E-06	2.03E-05	Alt	AFR
8	10683929	rs11776767	G	C	0.89	0.11	0.75	0.25	16.73	4.31E-05	2.29E-04	Ref	NAT
8	18257854	rs1801280	T	C	0.72	0.28	0.89	0.11	46.46	9.37E-12	1.14E-10	Ref	NAT
8	19819724	rs328	C	G	0.98	0.02	0.89	0.11	13.33	2.61E-04	1.21E-03	Ref	AFR
8	19832646	rs17482753	G	T	0.98	0.02	0.90	0.10	12.18	4.83E-04	2.13E-03	Ref	AFR
8	19847690	rs10503669	C	A	0.98	0.02	0.90	0.10	12.18	4.83E-04	2.13E-03	Ref	AFR
8	23059324	rs20575	C	G	0.21	0.79	0.13	0.87	8.82	2.97E-03	1.12E-02	Alt	NAT
8	37096059	rs1015003	G	T	0.57	0.43	0.31	0.69	56.10	6.88E-14	1.02E-12	Ref	EUR
8	57095808	rs10958476	T	C	0.92	0.08	0.85	0.15	7.67	5.60E-03	1.99E-02	Ref	AFR
8	65999963	rs2980003	C	T	0.46	0.54	0.64	0.36	24.57	7.16E-07	4.92E-06	Alt	NAT
8	69389217	rs1517114	C	G	0.26	0.74	0.13	0.87	25.27	4.98E-07	3.53E-06	Alt	NAT
8	89993488	rs10429371	C	T	0.38	0.62	0.27	0.73	10.76	1.04E-03	4.32E-03	Alt	EUR
8	103200036	rs1265138	A	G	0.92	0.08	0.77	0.23	20.80	5.11E-06	3.12E-05	Ref	EUR
8	118184783	rs13266634	C	T	0.74	0.26	0.58	0.42	16.35	5.27E-05	2.75E-04	Ref	AFR
8	122275906	rs7834765	G	T	0.28	0.72	0.53	0.47	43.72	3.78E-11	4.36E-10	Alt	NAT

Table 5 (continued).

8	122909687	rs956225	A	G	0.97	0.03	0.87	0.13	15.09	1.03E-04	5.09E-04	Ref	AFR
8	126486409	rs17321515	A	G	0.52	0.48	0.43	0.57	6.96	8.32E-03	2.85E-02	Ref	EUR
8	126490972	rs2954029	A	T	0.54	0.46	0.43	0.57	7.75	5.36E-03	1.91E-02	Ref	AFR
8	129787976	rs10956445	T	C	0.36	0.64	0.47	0.53	7.65	5.68E-03	2.01E-02	Alt	NAT
8	132330716	rs10108033	T	C	0.20	0.80	0.44	0.56	38.29	6.10E-10	6.16E-09	Alt	AFR
8	139884509	rs6988229	C	T	0.81	0.19	0.96	0.04	125.95	3.16E-29	1.20E-27	Ref	NAT
8	143999600	rs1799998	A	G	0.43	0.57	0.71	0.29	62.59	2.55E-15	4.27E-14	Alt	EUR
9	4744743	rs409801	T	C	0.74	0.26	0.41	0.59	76.50	2.20E-18	4.64E-17	Ref	AFR
9	4763176	rs385893	T	C	0.55	0.45	0.25	0.75	85.51	2.31E-20	5.52E-19	Ref	AFR
9	4814948	rs13300663	G	C	0.78	0.22	0.60	0.40	23.55	1.21E-06	8.12E-06	Ref	AFR
9	21499624	rs7849420	A	C	0.39	0.61	0.14	0.86	89.71	2.76E-21	6.99E-20	Alt	NAT
9	35141705	rs10972341	A	G	0.22	0.78	0.66	0.34	147.21	7.07E-34	3.28E-32	Alt	AFR
9	35648950	rs3138083	A	G	0.55	0.45	0.20	0.80	127.98	1.13E-29	4.41E-28	Ref	AFR
9	86920236	rs7867504	T	C	0.25	0.75	0.40	0.60	16.49	4.89E-05	2.58E-04	Alt	AFR
9	93636664	rs290227	G	A	0.83	0.17	0.45	0.55	97.24	6.15E-23	1.70E-21	Ref	EUR
9	104223233	rs10819937	C	G	0.51	0.49	0.34	0.66	23.75	1.09E-06	7.34E-06	Ref	NAT
9	104378003	rs10121600	C	T	0.68	0.32	0.77	0.23	9.79	1.75E-03	6.99E-03	Ref	NAT
9	104867838	rs10989824	C	T	0.66	0.34	0.47	0.53	25.71	3.96E-07	2.85E-06	Ref	EUR
9	107562804	rs2230808	T	C	0.15	0.85	0.38	0.62	37.89	7.50E-10	7.48E-09	Alt	EUR
9	109632353	rs4743034	G	A	0.92	0.08	0.81	0.19	12.47	4.13E-04	1.85E-03	Ref	NAT
9	111455575	rs10512385	A	G	0.96	0.04	0.90	0.10	6.54	1.06E-02	3.55E-02	Ref	NAT
9	112521126	rs4978848	G	C	0.20	0.80	0.34	0.66	13.96	1.87E-04	8.87E-04	Alt	NAT
9	114293634	rs10980926	A	G	0.79	0.21	0.55	0.45	38.07	6.81E-10	6.84E-09	Ref	AFR
9	119249339	rs7852872	C	G	0.36	0.64	0.54	0.46	21.28	3.96E-06	2.46E-05	Alt	NAT
9	120472764	rs1927907	C	T	0.93	0.07	0.75	0.25	28.46	9.57E-08	7.41E-07	Ref	EUR

Table 5 (continued).

9	124565820	rs10760187	T	C	0.55	0.45	0.25	0.75	85.51	2.31E-20	5.52E-19	Ref	EUR
9	125137695	rs10306135	A	T	0.92	0.08	0.97	0.03	25.60	4.20E-07	2.99E-06	Ref	NAT
9	128001119	rs430397	C	T	0.98	0.02	0.86	0.14	19.41	1.06E-05	6.16E-05	Ref	EUR
9	132501881	rs2302821	A	C	0.76	0.24	0.51	0.49	43.53	4.18E-11	4.78E-10	Ref	EUR
10	26734587	rs2992257	C	T	0.91	0.09	0.54	0.46	91.06	1.39E-21	3.59E-20	Ref	AFR
10	30834632	rs11008099	G	A	0.98	0.02	0.82	0.18	30.93	2.67E-08	2.22E-07	Ref	AFR
10	36885921	rs1360573	A	G	0.95	0.05	0.88	0.12	7.82	5.16E-03	1.84E-02	Ref	NAT
10	52010708	rs10508921	C	T	0.95	0.05	0.70	0.30	51.79	6.17E-13	8.46E-12	Ref	EUR
10	64963449	rs4379723	T	C	0.81	0.19	0.62	0.38	27.13	1.91E-07	1.42E-06	Ref	AFR
10	65027610	rs10761731	A	T	0.82	0.18	0.66	0.34	19.08	1.25E-05	7.26E-05	Ref	AFR
10	65104500	rs7896518	A	G	0.82	0.18	0.66	0.34	19.08	1.25E-05	7.26E-05	Ref	AFR
10	65133822	rs7923609	A	G	0.81	0.19	0.65	0.35	20.19	7.00E-06	4.21E-05	Ref	AFR
10	65658744	rs938036	T	C	0.01	0.99	0.06	0.94	8.61	3.35E-03	1.25E-02	Alt	EUR
10	68598292	rs7902091	C	A	0.39	0.61	0.66	0.34	52.99	3.35E-13	4.71E-12	Alt	AFR
10	73769507	rs4148943	C	T	0.59	0.41	0.82	0.18	56.97	4.43E-14	6.71E-13	Ref	NAT
10	73772762	rs730720	C	T	0.59	0.41	0.82	0.18	56.97	4.43E-14	6.71E-13	Ref	NAT
10	79211262	rs603788	G	C	0.46	0.54	0.68	0.32	34.83	3.59E-09	3.32E-08	Alt	EUR
10	79680434	rs754466	A	T	0.91	0.09	0.77	0.23	19.17	1.20E-05	6.95E-05	Ref	AFR
10	80960828	rs2802369	C	A	0.72	0.28	0.62	0.38	7.20	7.30E-03	2.52E-02	Ref	AFR
10	90826779	rs1937332	A	G	0.54	0.46	0.73	0.27	30.11	4.08E-08	3.31E-07	Ref	NAT
10	94839642	rs2068888	G	A	0.51	0.49	0.25	0.75	64.03	1.22E-15	2.10E-14	Ref	AFR
10	96541616	rs4244285	G	A	0.94	0.06	0.70	0.30	47.09	6.79E-12	8.39E-11	Ref	EUR
10	96707202	rs4086116	C	T	0.96	0.04	0.89	0.11	8.95	2.78E-03	1.05E-02	Ref	NAT
10	96725535	rs4917639	A	C	0.96	0.04	0.89	0.11	8.95	2.78E-03	1.05E-02	Ref	NAT
10	96798548	rs1934951	C	T	0.94	0.06	0.63	0.37	68.19	1.48E-16	2.75E-15	Ref	EUR

Table 5 (continued).

10	101542578	rs717620	C	T	0.92	0.08	0.78	0.22	19.52	9.95E-06	5.83E-05	Ref	AFR
10	101605693	rs3740065	A	G	0.91	0.09	0.72	0.28	30.11	4.08E-08	3.31E-07	Ref	EUR
10	101620771	rs12762549	C	G	0.68	0.32	0.44	0.56	38.18	6.47E-10	6.52E-09	Ref	AFR
10	101795361	rs10883437	T	A	0.59	0.41	0.71	0.29	11.47	7.08E-04	3.03E-03	Ref	NAT
10	112836503	rs1800544	G	C	0.45	0.55	0.58	0.42	11.71	6.23E-04	2.70E-03	Alt	EUR
10	112837538	rs1800545	G	A	0.96	0.04	0.82	0.18	21.41	3.70E-06	2.32E-05	Ref	EUR
10	115805056	rs1801253	G	C	0.11	0.89	0.23	0.77	14.67	1.28E-04	6.23E-04	Alt	NAT
10	119210375	rs6585436	T	C	0.99	0.01	0.95	0.05	7.51	6.14E-03	2.15E-02	Ref	EUR
10	122900623	rs2901286	C	A	0.92	0.08	0.85	0.15	6.54	1.06E-02	3.55E-02	Ref	AFR
10	124214448	rs10490924	G	T	0.76	0.24	0.62	0.38	14.43	1.45E-04	7.03E-04	Ref	EUR
10	124219275	rs3793917	C	G	0.76	0.24	0.61	0.39	15.57	7.96E-05	4.02E-04	Ref	EUR
11	243268	rs505404	T	G	0.73	0.27	0.95	0.05	160.61	8.31E-37	4.25E-35	Ref	NAT
11	4159457	rs9937	A	G	0.70	0.30	0.60	0.40	7.08	7.78E-03	2.68E-02	Ref	AFR
11	8639200	rs4929923	T	C	0.32	0.68	0.57	0.43	44.39	2.69E-11	3.15E-10	Alt	EUR
11	12159661	rs1994318	C	A	0.74	0.26	0.50	0.50	39.55	3.19E-10	3.32E-09	Ref	AFR
11	13293905	rs900145	C	T	0.52	0.48	0.43	0.57	6.15	1.32E-02	4.31E-02	Ref	AFR
11	20659757	rs2298826	G	A	0.64	0.36	0.52	0.48	9.43	2.13E-03	8.37E-03	Ref	AFR
11	27667202	rs925946	T	G	0.17	0.83	0.08	0.92	17.51	2.85E-05	1.56E-04	Alt	NAT
11	27670108	rs10501087	T	C	0.90	0.10	0.58	0.42	70.52	4.54E-17	8.72E-16	Ref	AFR
11	27677041	rs7124442	C	T	0.19	0.81	0.10	0.90	12.18	4.83E-04	2.13E-03	Alt	NAT
11	27679916	rs6265	C	T	0.90	0.10	0.56	0.44	80.26	3.28E-19	7.25E-18	Ref	AFR
11	27684517	rs11030104	A	G	0.90	0.10	0.55	0.45	82.83	8.92E-20	2.06E-18	Ref	AFR
11	27700125	rs7103411	C	T	0.10	0.90	0.45	0.55	82.83	8.92E-20	2.06E-18	Alt	AFR
11	32895664	rs10767971	T	C	0.18	0.82	0.52	0.48	76.54	2.16E-18	4.54E-17	Alt	AFR
11	35123051	rs1559759	C	A	0.93	0.07	0.70	0.30	42.61	6.70E-11	7.51E-10	Ref	AFR

Table 5 (continued).

11	43648368	rs4237643	T	G	0.12	0.88	0.20	0.80	7.21	7.27E-03	2.51E-02	Alt	NAT
11	61557803	rs102275	T	C	0.16	0.84	0.65	0.35	173.19	1.48E-39	8.48E-38	Alt	AFR
11	64048912	rs477895	C	T	0.08	0.92	0.26	0.74	27.60	1.49E-07	1.14E-06	Alt	EUR
11	64360274	rs11231825	T	C	0.78	0.22	0.66	0.34	10.56	1.16E-03	4.78E-03	Ref	AFR
11	69462910	rs9344	G	A	0.70	0.30	0.45	0.55	44.00	3.29E-11	3.81E-10	Ref	AFR
11	74907582	rs2306168	C	T	0.90	0.10	0.79	0.21	11.66	6.40E-04	2.76E-03	Ref	EUR
11	80377052	rs17140547	C	T	0.99	0.01	0.80	0.20	39.18	3.86E-10	3.98E-09	Ref	AFR
11	81235150	rs2032381	G	T	0.98	0.02	0.71	0.29	62.59	2.55E-15	4.27E-14	Ref	AFR
11	82564294	rs2229437	T	G	0.95	0.05	0.88	0.12	7.82	5.16E-03	1.84E-02	Ref	AFR
11	101771433	rs11225055	T	C	0.89	0.11	0.70	0.30	32.03	1.52E-08	1.30E-07	Ref	AFR
11	112026156	rs5744247	G	C	0.78	0.22	0.59	0.41	24.87	6.14E-07	4.27E-06	Ref	AFR
11	113306765	rs4436578	C	T	0.28	0.72	0.37	0.63	6.46	1.11E-02	3.70E-02	Alt	EUR
11	119099906	rs4938642	G	C	0.99	0.01	0.82	0.18	35.51	2.54E-09	2.38E-08	Ref	AFR
11	131807171	rs12098973	A	G	0.92	0.08	0.75	0.25	28.02	1.20E-07	9.22E-07	Ref	EUR
11	133120937	rs2078454	C	A	0.84	0.16	0.91	0.09	10.53	1.18E-03	4.84E-03	Ref	NAT
12	2757769	rs2239128	T	C	0.05	0.95	0.36	0.64	69.14	9.19E-17	1.73E-15	Alt	EUR
12	6291093	rs7342306	G	A	0.84	0.16	0.63	0.37	30.89	2.73E-08	2.26E-07	Ref	AFR
12	6451590	rs4149570	A	C	0.29	0.71	0.45	0.55	17.35	3.11E-05	1.69E-04	Alt	AFR
12	6953257	rs11064426	A	C	0.69	0.31	0.52	0.48	18.49	1.71E-05	9.69E-05	Ref	EUR
12	10170727	rs11053548	A	G	0.92	0.08	0.77	0.23	22.49	2.11E-06	1.36E-05	Ref	AFR
12	10471050	rs12303914	A	G	0.09	0.91	0.26	0.74	24.04	9.43E-07	6.38E-06	Alt	NAT
12	11547532	rs2900174	A	G	0.84	0.16	0.91	0.09	9.94	1.62E-03	6.52E-03	Ref	EUR
12	11855624	rs2187642	A	C	0.18	0.82	0.54	0.46	91.26	1.26E-21	3.26E-20	Alt	EUR
12	20860093	rs3794271	G	A	0.74	0.26	0.27	0.73	186.00	2.37E-42	1.46E-40	Ref	AFR
12	21283322	rs4149015	G	A	0.99	0.01	0.89	0.11	19.20	1.18E-05	6.84E-05	Ref	AFR

Table 5 (continued).

12	21327740	rs4149036	C	A	0.81	0.19	0.55	0.45	45.96	1.21E-11	1.45E-10	Ref	EUR
12	21329738	rs2306283	A	G	0.53	0.47	0.29	0.71	48.20	3.85E-12	4.88E-11	Ref	EUR
12	21377559	rs4149080	G	C	0.81	0.19	0.61	0.39	28.63	8.76E-08	6.81E-07	Ref	EUR
12	21378021	rs4149081	G	A	0.81	0.19	0.61	0.39	28.63	8.76E-08	6.81E-07	Ref	EUR
12	25484193	rs7965364	C	T	0.36	0.64	0.27	0.73	6.62	1.01E-02	3.40E-02	Alt	EUR
12	50247468	rs7138803	G	A	0.80	0.20	0.69	0.31	8.98	2.73E-03	1.04E-02	Ref	AFR
12	50350953	rs296766	T	C	0.04	0.96	0.14	0.86	14.53	1.38E-04	6.70E-04	Alt	EUR
12	51357542	rs12304921	A	G	0.87	0.13	0.59	0.41	55.95	7.42E-14	1.10E-12	Ref	AFR
12	54270228	rs2120991	C	A	0.94	0.06	0.82	0.18	16.19	5.73E-05	2.98E-04	Ref	AFR
12	54736470	rs4326844	A	G	0.11	0.89	0.46	0.54	82.47	1.08E-19	2.47E-18	Alt	AFR
12	58144665	rs2069502	C	T	0.63	0.37	0.44	0.56	26.06	3.31E-07	2.40E-06	Ref	AFR
12	58145156	rs2270777	C	T	0.88	0.12	0.79	0.21	8.10	4.44E-03	1.60E-02	Ref	NAT
12	66359752	rs8756	C	A	0.33	0.67	0.17	0.83	30.31	3.69E-08	3.01E-07	Alt	NAT
12	77738005	rs6538140	G	A	0.26	0.74	0.43	0.57	18.82	1.43E-05	8.22E-05	Alt	AFR
12	88890671	rs995030	A	G	0.08	0.92	0.29	0.71	38.79	4.72E-10	4.83E-09	Alt	EUR
12	88953959	rs4474514	G	A	0.08	0.92	0.30	0.70	40.01	2.53E-10	2.66E-09	Alt	EUR
12	96479267	rs2660869	G	C	0.88	0.12	0.96	0.04	33.86	5.92E-09	5.29E-08	Ref	NAT
12	102965329	rs7964748	A	G	0.89	0.11	0.81	0.19	7.55	6.02E-03	2.12E-02	Ref	NAT
12	112190438	rs6490294	C	A	0.78	0.22	0.24	0.76	272.05	4.05E-61	4.04E-59	Ref	EUR
12	112817783	rs11066280	T	A	0.99	0.01	0.85	0.15	28.38	9.98E-08	7.72E-07	Ref	AFR
12	115094260	rs11067228	A	G	0.32	0.68	0.54	0.46	30.65	3.10E-08	2.55E-07	Alt	EUR
12	117327592	rs7294919	T	C	0.93	0.07	0.77	0.23	24.26	8.43E-07	5.74E-06	Ref	EUR
12	124460167	rs4765127	G	T	0.74	0.26	0.84	0.16	14.27	1.59E-04	7.64E-04	Ref	NAT
12	125312425	rs10846744	G	C	0.81	0.19	0.48	0.52	73.88	8.30E-18	1.67E-16	Ref	EUR
12	125348263	rs4238001	C	T	0.94	0.06	0.97	0.03	9.22	2.40E-03	9.32E-03	Ref	NAT

Table 5 (continued).

12	129637348	rs12824981	C	T	0.81	0.19	0.88	0.12	6.57	1.03E-02	3.48E-02	Ref	NAT
12	133417802	rs3741489	T	C	0.75	0.25	0.66	0.34	5.89	1.52E-02	4.94E-02	Ref	EUR
13	21876096	rs9788333	C	G	0.86	0.14	0.70	0.30	20.42	6.22E-06	3.75E-05	Ref	AFR
13	27415673	rs9319321	A	T	0.53	0.47	0.32	0.68	35.17	3.02E-09	2.80E-08	Ref	NAT
13	46703142	rs7324845	G	A	0.06	0.94	0.13	0.87	8.50	3.56E-03	1.32E-02	Alt	EUR
13	47409034	rs6314	G	A	0.94	0.06	0.97	0.03	9.22	2.40E-03	9.32E-03	Ref	NAT
13	47469940	rs6313	G	A	0.74	0.26	0.50	0.50	35.79	2.19E-09	2.07E-08	Ref	AFR
13	52566126	rs9535826	T	G	0.65	0.35	0.47	0.53	21.25	4.03E-06	2.50E-05	Ref	AFR
13	95863008	rs11568658	C	A	0.92	0.08	0.85	0.15	6.75	9.36E-03	3.17E-02	Ref	AFR
13	103528002	rs17655	G	C	0.72	0.28	0.57	0.43	15.00	1.07E-04	5.30E-04	Ref	EUR
14	23861811	rs365990	A	G	0.88	0.12	0.77	0.23	13.08	2.99E-04	1.37E-03	Ref	NAT
14	23865885	rs452036	G	A	0.88	0.12	0.78	0.22	10.98	9.21E-04	3.87E-03	Ref	NAT
14	41523462	rs1959947	A	G	0.44	0.56	0.26	0.74	29.47	5.69E-08	4.54E-07	Alt	NAT
14	48015982	rs1160351	A	C	0.75	0.25	0.40	0.60	91.66	1.03E-21	2.67E-20	Ref	AFR
14	81598912	rs17111530	T	C	0.96	0.04	0.74	0.26	41.98	9.23E-11	1.02E-09	Ref	EUR
14	97171075	rs17244419	C	T	0.98	0.02	0.82	0.18	29.51	5.57E-08	4.45E-07	Ref	AFR
14	101159416	rs7149242	T	G	0.25	0.75	0.41	0.59	16.49	4.89E-05	2.58E-04	Alt	EUR
14	101679255	rs6575836	G	A	0.29	0.71	0.41	0.59	9.71	1.83E-03	7.27E-03	Alt	AFR
14	103040087	rs11628318	T	A	0.58	0.42	0.40	0.60	23.55	1.21E-06	8.12E-06	Ref	AFR
14	103566785	rs2297067	C	T	0.87	0.13	0.74	0.26	15.99	6.38E-05	3.27E-04	Ref	AFR
14	104165753	rs861539	G	A	0.94	0.06	0.87	0.13	7.52	6.11E-03	2.15E-02	Ref	NAT
14	105263608	rs2494752	A	G	0.38	0.62	0.64	0.36	49.15	2.37E-12	3.06E-11	Alt	EUR
15	28365618	rs12913832	A	G	0.89	0.11	0.94	0.06	8.61	3.35E-03	1.25E-02	Ref	AFR
15	29731444	rs11856574	G	A	0.80	0.20	0.67	0.33	13.96	1.87E-04	8.87E-04	Ref	EUR
15	42683787	rs2412710	G	A	0.96	0.04	0.99	0.01	8.10	4.44E-03	1.60E-02	Ref	NAT

Table 5 (continued).

15	45500047	rs765787	A	G	0.82	0.18	0.70	0.30	11.20	8.16E-04	3.46E-03	Ref	AFR
15	48392165	rs1834640	A	G	0.45	0.55	0.26	0.74	31.40	2.10E-08	1.77E-07	Alt	AFR
15	51545454	rs12907866	A	G	0.79	0.21	0.54	0.46	43.80	3.63E-11	4.19E-10	Ref	AFR
15	54380200	rs11071033	T	C	0.76	0.24	0.58	0.42	21.77	3.08E-06	1.95E-05	Ref	AFR
15	58683366	rs1532085	A	G	0.36	0.64	0.46	0.54	6.06	1.38E-02	4.50E-02	Alt	EUR
15	60781513	rs3743266	T	C	0.88	0.12	0.75	0.25	15.31	9.15E-05	4.58E-04	Ref	NAT
15	65183801	rs1719271	A	G	0.59	0.41	0.39	0.61	26.83	2.23E-07	1.65E-06	Ref	EUR
15	71424009	rs12904863	T	C	0.94	0.06	0.69	0.31	46.57	8.83E-12	1.08E-10	Ref	AFR
15	74176557	rs16958445	G	A	0.99	0.01	0.88	0.12	18.36	1.83E-05	1.03E-04	Ref	AFR
15	75041341	rs2069526	T	G	0.97	0.03	0.92	0.08	6.30	1.20E-02	3.98E-02	Ref	EUR
15	75043281	rs4646425	C	T	0.97	0.03	0.92	0.08	6.30	1.20E-02	3.98E-02	Ref	AFR
15	75045692	rs4646427	T	C	0.97	0.03	0.92	0.08	6.30	1.20E-02	3.98E-02	Ref	EUR
15	75047426	rs2470890	C	T	0.88	0.12	0.80	0.20	6.35	1.17E-02	3.89E-02	Ref	AFR
15	78865425	rs588765	T	C	0.09	0.91	0.20	0.80	11.91	5.58E-04	2.43E-03	Alt	NAT
15	79552379	rs8025118	G	T	0.71	0.29	0.54	0.46	19.89	8.21E-06	4.87E-05	Ref	EUR
15	87064089	rs12591257	A	C	0.99	0.01	0.92	0.08	11.99	5.34E-04	2.33E-03	Ref	AFR
16	14388305	rs1659127	G	A	0.86	0.14	0.52	0.48	79.33	5.26E-19	1.15E-17	Ref	AFR
16	16236650	rs212091	T	C	0.91	0.09	0.78	0.22	17.93	2.29E-05	1.27E-04	Ref	AFR
16	23909952	rs9922316	T	G	0.10	0.90	0.05	0.95	10.63	1.12E-03	4.62E-03	Alt	NAT
16	27375787	rs8832	A	G	0.38	0.62	0.56	0.44	21.47	3.59E-06	2.25E-05	Alt	EUR
16	27376217	rs1029489	A	G	0.35	0.65	0.55	0.45	27.51	1.56E-07	1.18E-06	Alt	EUR
16	30918487	rs11649653	C	G	0.61	0.39	0.16	0.84	261.05	1.01E-58	9.51E-57	Ref	AFR
16	31048079	rs10871454	C	T	0.61	0.39	0.17	0.83	246.96	1.19E-55	1.05E-53	Ref	AFR
16	31099011	rs11150606	T	C	0.65	0.35	0.33	0.67	76.96	1.74E-18	3.69E-17	Ref	AFR
16	31102321	rs7294	C	T	0.44	0.56	0.89	0.11	351.32	2.19E-78	3.01E-76	Alt	AFR

Table 5 (continued).

16	31103796	rs2359612	A	G	0.39	0.61	0.83	0.17	246.96	1.19E-55	1.05E-53	Alt	AFR
16	31104509	rs8050894	C	G	0.61	0.39	0.16	0.84	246.96	1.19E-55	1.05E-53	Ref	AFR
16	31104878	rs9934438	G	A	0.61	0.39	0.17	0.83	233.86	8.59E-53	7.07E-51	Ref	AFR
16	31107689	rs9923231	C	T	0.62	0.38	0.17	0.83	240.14	3.67E-54	3.11E-52	Ref	AFR
16	31110981	rs7196161	G	A	0.39	0.61	0.84	0.16	261.05	1.01E-58	9.51E-57	Alt	EUR
16	53769677	rs6499640	G	A	0.83	0.17	0.74	0.26	6.90	8.62E-03	2.95E-02	Ref	NAT
16	53816275	rs8050136	C	A	0.92	0.08	0.80	0.20	13.57	2.29E-04	1.08E-03	Ref	NAT
16	53820527	rs9939609	T	A	0.92	0.08	0.80	0.20	14.71	1.26E-04	6.13E-04	Ref	NAT
16	53876751	rs12595985	C	A	0.94	0.06	0.74	0.26	34.94	3.41E-09	3.15E-08	Ref	EUR
16	55844609	rs2244613	G	T	0.36	0.64	0.54	0.46	21.28	3.96E-06	2.46E-05	Alt	EUR
16	58567238	rs37062	A	G	0.48	0.52	0.66	0.34	27.03	2.01E-07	1.49E-06	Alt	NAT
16	65895364	rs1126179	A	G	0.35	0.65	0.22	0.78	16.72	4.33E-05	2.30E-04	Alt	NAT
16	72108093	rs2000999	G	A	0.89	0.11	0.71	0.29	27.55	1.53E-07	1.16E-06	Ref	AFR
16	75167579	rs7195303	G	A	0.07	0.93	0.20	0.80	17.79	2.46E-05	1.36E-04	Alt	AFR
16	80497601	rs4581712	C	A	0.81	0.19	0.57	0.43	36.65	1.42E-09	1.37E-08	Ref	AFR
16	84046715	rs11864146	A	G	0.81	0.19	0.93	0.07	39.54	3.21E-10	3.33E-09	Ref	EUR
16	84987679	rs2326458	C	A	0.55	0.45	0.34	0.66	33.92	5.75E-09	5.16E-08	Ref	AFR
16	85906616	rs11648716	A	G	0.96	0.04	0.99	0.01	12.65	3.76E-04	1.69E-03	Ref	NAT
16	85961562	rs17444745	G	A	0.92	0.08	0.83	0.17	9.62	1.92E-03	7.63E-03	Ref	AFR
16	89613123	rs2292954	A	G	0.97	0.03	0.86	0.14	16.29	5.43E-05	2.83E-04	Ref	AFR
16	89985940	rs2228479	G	A	0.99	0.01	0.82	0.18	33.18	8.40E-09	7.41E-08	Ref	AFR
16	89986608	rs2228478	A	G	0.96	0.04	0.80	0.20	28.82	7.93E-08	6.19E-07	Ref	EUR
17	7091650	rs314253	T	C	0.42	0.58	0.54	0.46	9.46	2.10E-03	8.26E-03	Alt	NAT
17	18232096	rs1979277	G	A	0.78	0.22	0.89	0.11	21.39	3.75E-06	2.34E-05	Ref	NAT
17	19804247	rs397969	T	C	0.62	0.38	0.79	0.21	27.63	1.47E-07	1.12E-06	Ref	NAT

Table 5 (continued).

17	28525011	rs1042173	A	C	0.34	0.66	0.25	0.75	7.11	7.64E-03	2.63E-02	Alt	NAT
17	28564346	rs25531	T	C	0.97	0.03	0.90	0.10	8.35	3.86E-03	1.42E-02	Ref	EUR
17	32574448	rs4795893	G	A	0.30	0.70	0.40	0.60	6.31	1.20E-02	3.98E-02	Alt	AFR
17	32583132	rs2857657	G	C	0.02	0.98	0.09	0.91	10.53	1.18E-03	4.84E-03	Alt	AFR
17	32583269	rs4586	T	C	0.31	0.69	0.40	0.60	7.05	7.93E-03	2.73E-02	Alt	AFR
17	43892520	rs242941	A	C	0.25	0.75	0.13	0.87	20.88	4.88E-06	2.99E-05	Alt	NAT
17	49960309	rs967676	T	C	0.85	0.15	0.73	0.27	12.97	3.17E-04	1.45E-03	Ref	AFR
17	54850329	rs4794665	A	G	0.11	0.89	0.19	0.81	6.51	1.08E-02	3.60E-02	Alt	NAT
17	61712964	rs7209435	T	C	0.76	0.24	0.88	0.12	22.67	1.93E-06	1.25E-05	Ref	NAT
17	68259446	rs623011	A	G	0.24	0.76	0.42	0.58	21.77	3.08E-06	1.95E-05	Alt	AFR
17	68326338	rs312691	T	C	0.77	0.23	0.60	0.40	19.31	1.11E-05	6.46E-05	Ref	AFR
17	68976415	rs6501431	C	T	0.14	0.86	0.05	0.95	23.00	1.62E-06	1.07E-05	Alt	NAT
17	70098161	rs9913711	G	C	0.08	0.92	0.23	0.77	20.80	5.11E-06	3.12E-05	Alt	NAT
17	74499400	rs12948783	G	A	0.85	0.15	0.92	0.08	11.99	5.34E-04	2.33E-03	Ref	NAT
17	75313335	rs4788985	A	G	0.36	0.64	0.48	0.52	9.42	2.14E-03	8.41E-03	Alt	NAT
17	78324259	rs12051723	G	T	0.96	0.04	0.76	0.24	39.37	3.50E-10	3.62E-09	Ref	AFR
17	78348494	rs6565681	A	G	0.63	0.37	0.37	0.63	48.82	2.80E-12	3.59E-11	Ref	NAT
18	3512216	rs17724172	T	C	0.94	0.06	0.83	0.17	15.01	1.07E-04	5.29E-04	Ref	AFR
18	5978931	rs1539808	C	T	0.98	0.02	0.80	0.20	31.62	1.87E-08	1.58E-07	Ref	EUR
18	20720973	rs11082304	G	T	0.69	0.31	0.53	0.47	18.51	1.69E-05	9.58E-05	Ref	AFR
18	22044400	rs4483927	G	T	0.20	0.80	0.37	0.63	21.21	4.12E-06	2.55E-05	Alt	EUR
18	55816791	rs4149601	G	A	0.94	0.06	0.75	0.25	31.88	1.64E-08	1.40E-07	Ref	NAT
18	57673799	rs12964056	A	G	0.37	0.63	0.28	0.72	7.53	6.07E-03	2.14E-02	Alt	EUR
18	61176266	rs2032224	C	A	0.39	0.61	0.22	0.78	28.50	9.36E-08	7.27E-07	Alt	NAT
18	62898666	rs637644	A	G	0.31	0.69	0.20	0.80	13.27	2.69E-04	1.25E-03	Alt	NAT

Table 5 (continued).

18	63031236	rs470490	G	T	0.21	0.79	0.37	0.63	18.51	1.69E-05	9.59E-05	Alt	AFR
19	8672000	rs7249094	G	A	0.46	0.54	0.21	0.79	65.16	6.91E-16	1.21E-14	Alt	NAT
19	10000322	rs1862471	C	G	0.44	0.56	0.65	0.35	33.38	7.57E-09	6.70E-08	Alt	EUR
19	10310392	rs2116940	A	G	0.74	0.26	0.63	0.37	9.10	2.55E-03	9.86E-03	Ref	EUR
19	10395683	rs5498	A	G	0.28	0.72	0.70	0.30	145.21	1.93E-33	8.77E-32	Alt	EUR
19	10397403	rs3093030	C	T	0.28	0.72	0.75	0.25	194.28	3.70E-44	2.39E-42	Alt	EUR
19	11227326	rs2738446	C	G	0.49	0.51	0.79	0.21	88.10	6.22E-21	1.54E-19	Alt	EUR
19	15990431	rs2108622	C	T	0.88	0.12	0.76	0.24	14.17	1.67E-04	7.99E-04	Ref	AFR
19	29736342	rs11083866	G	A	0.69	0.31	0.55	0.45	13.71	2.14E-04	1.01E-03	Ref	AFR
19	39735106	rs8103142	T	C	0.62	0.38	0.88	0.12	109.71	1.14E-25	3.65E-24	Ref	NAT
19	39738787	rs12979860	C	T	0.62	0.38	0.88	0.12	109.71	1.14E-25	3.65E-24	Ref	NAT
19	41356379	rs28399433	A	C	0.91	0.09	0.77	0.23	16.40	5.12E-05	2.69E-04	Ref	EUR
19	41515702	rs2279345	T	C	0.10	0.90	0.32	0.68	37.15	1.09E-09	1.07E-08	Alt	AFR
19	41858876	rs1800471	C	G	0.98	0.02	0.99	0.01	9.05	2.62E-03	9.98E-03	Ref	NAT
19	41860296	rs1800469	A	G	0.57	0.43	0.45	0.55	9.50	2.06E-03	8.11E-03	Ref	NAT
19	45376284	rs519113	C	G	0.71	0.29	0.81	0.19	11.13	8.52E-04	3.60E-03	Ref	NAT
19	45403412	rs1160985	C	T	0.43	0.57	0.65	0.35	37.48	9.22E-10	9.09E-09	Alt	AFR
19	45412079	rs7412	C	T	0.99	0.01	0.90	0.10	16.73	4.30E-05	2.29E-04	Ref	EUR
19	45414451	rs439401	T	C	0.76	0.24	0.53	0.47	37.78	7.93E-10	7.89E-09	Ref	NAT
19	45415640	rs445925	G	A	0.98	0.02	0.88	0.12	14.51	1.40E-04	6.77E-04	Ref	EUR
19	46172278	rs11671664	G	A	0.96	0.04	0.60	0.40	91.20	1.30E-21	3.35E-20	Ref	AFR
20	17122593	rs852069	A	G	0.63	0.37	0.72	0.28	6.53	1.06E-02	3.55E-02	Ref	NAT
20	38820805	rs6028945	G	T	0.96	0.04	0.88	0.12	9.58	1.97E-03	7.79E-03	Ref	AFR
20	45288453	rs6066043	G	A	0.76	0.24	0.51	0.49	41.53	1.16E-10	1.27E-09	Ref	AFR
20	60791404	rs3787429	C	T	0.44	0.56	0.59	0.41	17.78	2.48E-05	1.37E-04	Alt	AFR

Table 5 (continued).

20	61981104	rs1044397	C	T	0.46	0.54	0.63	0.37	19.77	8.73E-06	5.15E-05	Alt	EUR
20	61981134	rs1044396	G	A	0.46	0.54	0.74	0.26	67.74	1.87E-16	3.44E-15	Alt	EUR
21	28217320	rs402007	C	G	0.65	0.35	0.50	0.50	14.71	1.26E-04	6.13E-04	Ref	NAT
21	41415044	rs9981861	T	C	0.70	0.30	0.79	0.21	7.93	4.86E-03	1.75E-02	Ref	NAT
21	42216955	rs2837857	C	T	0.69	0.31	0.77	0.23	6.52	1.07E-02	3.57E-02	Ref	EUR
21	46957794	rs1051266	T	C	0.37	0.63	0.49	0.51	9.42	2.15E-03	8.42E-03	Alt	NAT
22	19951207	rs4818	C	G	0.85	0.15	0.64	0.36	34.76	3.74E-09	3.44E-08	Ref	AFR
22	19952132	rs4646316	C	T	0.89	0.11	0.68	0.32	33.25	8.12E-09	7.17E-08	Ref	AFR
22	19956781	rs165599	G	A	0.65	0.35	0.44	0.56	31.01	2.56E-08	2.14E-07	Ref	AFR
22	22512415	rs987710	G	A	0.18	0.82	0.30	0.70	13.41	2.50E-04	1.17E-03	Alt	NAT
22	24295286	rs2739330	T	C	0.82	0.18	0.60	0.40	37.10	1.12E-09	1.10E-08	Ref	AFR
22	29196757	rs2269577	G	C	0.66	0.34	0.46	0.54	29.08	6.94E-08	5.49E-07	Ref	EUR
22	37097564	rs2284018	C	T	0.84	0.16	0.71	0.29	14.99	1.08E-04	5.35E-04	Ref	AFR
22	40075400	rs2294369	G	A	0.82	0.18	0.91	0.09	15.52	8.15E-05	4.10E-04	Ref	AFR
22	42524947	rs3892097	C	T	0.94	0.06	0.97	0.03	7.42	6.46E-03	2.26E-02	Ref	NAT
22	42526694	rs1065852	G	A	0.93	0.07	0.49	0.51	136.08	1.92E-31	8.05E-30	Ref	AFR
22	42528976	rs28360521	C	T	0.93	0.07	0.49	0.51	132.43	1.21E-30	4.87E-29	Ref	EUR
22	46594035	rs149711321	T	C	0.98	0.02	1.00	0.00	9.05	2.62E-03	9.98E-03	Ref	NAT
Puerto Rico													
1	16505320	rs1497406	A	G	0.59	0.41	0.46	0.54	14.10	1.73E-04	5.92E-03	Ref	AFR
1	65992625	rs1751492	C	T	0.20	0.80	0.37	0.63	26.72	2.35E-07	2.89E-05	Alt	EUR
1	76839536	rs12144344	C	T	0.80	0.20	0.67	0.33	15.81	7.00E-05	2.71E-03	Ref	AFR
1	154418879	rs4537545	C	T	0.45	0.55	0.55	0.45	9.41	2.15E-03	4.20E-02	Alt	AFR
1	226013355	rs1877724	C	T	0.85	0.15	0.76	0.24	9.50	2.05E-03	4.06E-02	Ref	AFR
2	10262920	rs1130609	T	G	0.40	0.60	0.29	0.71	12.27	4.60E-04	1.26E-02	Alt	AFR

Table 5 (continued).

2	27742603	rs780093	T	C	0.32	0.68	0.42	0.58	8.69	3.21E-03	4.97E-02	Alt	AFR
2	105897740	rs1020064	T	G	0.14	0.86	0.23	0.77	8.91	2.84E-03	4.51E-02	Alt	AFR
2	108994808	rs1402467	C	G	0.79	0.21	0.68	0.32	12.68	3.69E-04	1.08E-02	Ref	NAT
2	113594867	rs16944	A	G	0.53	0.47	0.38	0.62	19.61	9.48E-06	6.14E-04	Ref	AFR
2	113598107	rs4848306	G	A	0.66	0.34	0.54	0.46	12.11	5.02E-04	1.32E-02	Ref	AFR
2	130468366	rs10928927	C	T	0.31	0.69	0.21	0.79	12.93	3.24E-04	9.81E-03	Alt	EUR
2	152981335	rs16830728	G	T	0.91	0.09	0.81	0.19	11.62	6.51E-04	1.64E-02	Ref	EUR
2	201021954	rs1569175	T	C	0.19	0.81	0.11	0.89	16.47	4.94E-05	2.39E-03	Alt	EUR
3	2624938	rs2619566	G	A	0.36	0.64	0.20	0.80	32.49	1.20E-08	1.93E-06	Alt	EUR
3	45300605	rs33794	A	G	0.72	0.28	0.54	0.46	25.11	5.41E-07	5.03E-05	Ref	NAT
3	117574822	rs6438424	A	C	0.66	0.34	0.54	0.46	12.09	5.07E-04	1.33E-02	Ref	AFR
3	160429869	rs7624766	A	G	0.67	0.33	0.53	0.47	16.23	5.62E-05	2.39E-03	Ref	EUR
4	18017730	rs6830062	T	C	0.74	0.26	0.83	0.17	13.74	2.10E-04	6.88E-03	Ref	EUR
4	100239319	rs1229984	T	C	0.06	0.94	0.16	0.84	14.41	1.47E-04	5.15E-03	Alt	AFR
5	58713680	rs2547917	G	A	0.89	0.11	0.79	0.21	11.73	6.16E-04	1.56E-02	Ref	AFR
5	93810208	rs6869388	T	C	0.73	0.27	0.82	0.18	11.62	6.51E-04	1.64E-02	Ref	NAT
5	137707315	rs757647	G	A	0.82	0.18	0.66	0.34	24.56	7.22E-07	6.17E-05	Ref	EUR
5	165377699	rs9313296	G	C	0.86	0.14	0.92	0.08	9.22	2.39E-03	4.25E-02	Ref	NAT
6	7102084	rs675209	T	C	0.31	0.69	0.45	0.55	16.32	5.34E-05	2.39E-03	Alt	EUR
6	18139802	rs12201199	A	T	0.81	0.19	0.89	0.11	12.51	4.04E-04	1.14E-02	Ref	NAT
6	32448599	rs7748270	C	T	0.40	0.60	0.51	0.49	9.31	2.28E-03	4.25E-02	Alt	NAT
6	32604372	rs9272346	G	A	0.35	0.65	0.46	0.54	9.38	2.20E-03	4.25E-02	Alt	NAT
6	32972404	rs3128935	T	C	0.80	0.20	0.89	0.11	15.84	6.90E-05	2.67E-03	Ref	EUR
6	124951063	rs504008	A	C	0.44	0.56	0.33	0.67	11.56	6.75E-04	1.68E-02	Alt	EUR
6	142512136	rs225675	A	G	0.75	0.25	0.64	0.36	12.01	5.29E-04	1.38E-02	Ref	AFR

Table 5 (continued).

6	149329267	rs2500535	A	G	0.13	0.87	0.06	0.94	16.08	6.07E-05	2.39E-03	Alt	EUR
6	155929801	rs35229355	C	T	1.00	0.00	0.90	0.10	19.97	7.87E-06	5.21E-04	Ref	EUR
6	160113872	rs4880	A	G	0.46	0.54	0.61	0.39	22.12	2.56E-06	2.00E-04	Alt	EUR
6	160560845	rs628031	A	G	0.25	0.75	0.37	0.63	11.88	5.68E-04	1.46E-02	Alt	NAT
7	32444435	rs215738	G	A	0.16	0.84	0.09	0.91	13.03	3.06E-04	9.44E-03	Alt	NAT
7	37266352	rs4723619	T	C	0.88	0.12	0.95	0.05	16.22	5.64E-05	2.39E-03	Ref	EUR
7	80236014	rs13236689	T	G	0.67	0.33	0.56	0.44	11.23	8.06E-04	1.94E-02	Ref	EUR
7	87152103	rs4148740	A	G	0.92	0.08	0.85	0.15	9.45	2.11E-03	4.12E-02	Ref	NAT
7	87153585	rs10280101	A	C	0.92	0.08	0.85	0.15	9.45	2.11E-03	4.12E-02	Ref	NAT
7	87157051	rs7787082	G	A	0.80	0.20	0.68	0.32	13.76	2.08E-04	6.83E-03	Ref	EUR
7	87160561	rs2032583	A	G	0.92	0.08	0.85	0.15	9.45	2.11E-03	4.12E-02	Ref	NAT
7	87161049	rs4148739	T	C	0.92	0.08	0.85	0.15	9.45	2.11E-03	4.12E-02	Ref	NAT
7	87161520	rs11983225	T	C	0.92	0.08	0.85	0.15	9.45	2.11E-03	4.12E-02	Ref	NAT
7	87164986	rs10248420	A	G	0.82	0.18	0.69	0.31	14.23	1.62E-04	5.60E-03	Ref	EUR
7	94925820	rs854548	A	G	0.33	0.67	0.24	0.76	9.50	2.05E-03	4.06E-02	Alt	EUR
7	99361466	rs2242480	C	T	0.66	0.34	0.77	0.23	14.33	1.54E-04	5.35E-03	Ref	EUR
7	99382096	rs2740574	C	T	0.19	0.81	0.12	0.88	10.60	1.13E-03	2.55E-02	Alt	NAT
7	127164958	rs6467136	A	G	0.62	0.38	0.42	0.58	31.52	1.98E-08	2.99E-06	Ref	EUR
7	139702593	rs17837497	G	A	0.95	0.05	0.99	0.01	21.65	3.28E-06	2.44E-04	Ref	NAT
8	4078353	rs2407314	G	C	0.43	0.57	0.57	0.43	15.36	8.91E-05	3.35E-03	Alt	AFR
8	54119214	rs1425902	G	A	0.49	0.51	0.39	0.61	8.92	2.83E-03	4.50E-02	Alt	NAT
8	129787976	rs10956445	T	C	0.87	0.13	0.78	0.22	11.00	9.14E-04	2.15E-02	Ref	AFR
9	14446001	rs1556032	C	T	0.46	0.54	0.60	0.40	19.35	1.09E-05	6.94E-04	Alt	NAT
9	124565820	rs10760187	T	C	0.57	0.43	0.39	0.61	27.56	1.52E-07	1.98E-05	Ref	EUR
9	132501881	rs2302821	A	C	0.91	0.09	0.80	0.20	14.70	1.26E-04	4.48E-03	Ref	EUR

Table 5 (continued).

10	26734587	rs2992257	C	T	0.85	0.15	0.74	0.26	13.23	2.75E-04	8.70E-03	Ref	AFR
10	64963449	rs4379723	T	C	0.62	0.38	0.50	0.50	11.08	8.74E-04	2.07E-02	Ref	AFR
10	65133822	rs7923609	A	G	0.61	0.39	0.50	0.50	9.31	2.28E-03	4.25E-02	Ref	AFR
11	12159661	rs1994318	C	A	0.52	0.48	0.42	0.58	9.56	1.99E-03	3.94E-02	Ref	AFR
11	35123051	rs1559759	C	A	0.94	0.06	0.81	0.19	22.56	2.03E-06	1.61E-04	Ref	AFR
11	61557803	rs102275	T	C	0.39	0.61	0.63	0.37	49.51	1.97E-12	7.27E-10	Alt	AFR
11	67715028	rs308309	C	T	0.12	0.88	0.23	0.77	16.69	4.41E-05	2.24E-03	Alt	AFR
11	133120937	rs2078454	C	A	0.67	0.33	0.77	0.23	10.68	1.08E-03	2.47E-02	Ref	NAT
12	2757769	rs2239128	T	C	0.21	0.79	0.31	0.69	9.03	2.66E-03	4.28E-02	Alt	EUR
12	6291093	rs7342306	G	A	0.80	0.20	0.66	0.34	19.25	1.15E-05	7.30E-04	Ref	AFR
12	20860093	rs3794271	G	A	0.49	0.51	0.37	0.63	11.82	5.87E-04	1.50E-02	Alt	NAT
12	58144665	rs2069502	C	T	0.83	0.17	0.70	0.30	16.60	4.62E-05	2.29E-03	Ref	AFR
12	77436148	rs310786	C	T	0.13	0.88	0.21	0.79	9.34	2.24E-03	4.25E-02	Alt	AFR
12	115094260	rs11067228	A	G	0.45	0.55	0.57	0.43	11.28	7.83E-04	1.90E-02	Alt	EUR
12	117002658	rs2089222	G	A	0.81	0.19	0.90	0.10	22.13	2.55E-06	1.99E-04	Ref	EUR
13	52566126	rs9535826	T	G	0.63	0.37	0.45	0.55	28.08	1.16E-07	1.54E-05	Ref	AFR
14	73721775	rs11628713	C	T	0.89	0.11	0.79	0.21	13.95	1.88E-04	6.30E-03	Ref	NAT
14	101679255	rs6575836	G	A	0.14	0.86	0.24	0.76	9.64	1.91E-03	3.83E-02	Alt	AFR
14	105263608	rs2494752	A	G	0.13	0.87	0.22	0.78	10.08	1.50E-03	3.18E-02	Alt	EUR
15	63333724	rs3809566	A	G	0.21	0.79	0.31	0.69	9.03	2.66E-03	4.28E-02	Alt	EUR
15	71424009	rs12904863	T	C	0.95	0.05	0.88	0.12	10.23	1.38E-03	2.97E-02	Ref	AFR
15	78894339	rs1051730	G	A	0.77	0.23	0.64	0.36	14.18	1.66E-04	5.73E-03	Ref	NAT
16	7164219	rs1478693	T	G	0.55	0.45	0.65	0.35	10.28	1.34E-03	2.92E-02	Ref	EUR
16	10188060	rs7203315	G	T	0.83	0.17	0.90	0.10	10.38	1.27E-03	2.79E-02	Ref	NAT
16	14041958	rs1799801	T	C	0.81	0.19	0.68	0.32	15.00	1.07E-04	3.92E-03	Ref	AFR

Table 5 (continued).

16	31099011	rs11150606	T	C	0.94	0.06	0.86	0.14	9.29	2.31E-03	4.25E-02	Ref	AFR
16	51609947	rs2030114	G	A	0.68	0.32	0.80	0.20	17.18	3.39E-05	1.80E-03	Ref	NAT
16	84046715	rs11864146	A	G	0.83	0.17	0.90	0.10	14.16	1.68E-04	5.77E-03	Ref	EUR
17	49960309	rs967676	T	C	0.70	0.30	0.59	0.41	8.81	3.00E-03	4.73E-02	Ref	AFR
18	75605399	rs9961113	T	C	0.78	0.22	0.67	0.33	11.47	7.06E-04	1.75E-02	Ref	EUR
19	45403412	rs1160985	C	T	0.40	0.60	0.58	0.42	27.05	1.98E-07	2.48E-05	Alt	AFR
19	49206985	rs602662	G	A	0.50	0.50	0.61	0.39	10.65	1.10E-03	2.49E-02	Alt	AFR
20	61977556	rs2236196	G	A	0.44	0.56	0.31	0.69	16.31	5.37E-05	2.39E-03	Alt	NAT

[1]*Chromosome based on human reference genome hg19/GRCh37*
[2]*Position based on human reference genome hg19/GRCh37*
[3]*dbSNP ID of the SNP*
[4]*Base at the reference position*
[5]*Alternate base at the position*
[6]*Observed reference allele frequency*
[7]*Observed alternate allele frequency*
[8]*Expected reference allele frequency as computed from the ancestry proportions and the corresponding observed frequencies of the reference allele in the ancestral populations*
[9]*Expected alternate allele frequency as computed from the ancestry proportions and the corresponding observed frequencies of the alternate allele in the ancestral populations*
[10]*Chi-square statistic calculated based on* $\sum(Observed - Expected)^2$ *for reference and alternate alleles*
[11]*P-value calculated based off the chi-square distribution*
[12]*Corresponding FDR q-value*
[13]*Allele (Reference or Alternate) that is enriched in a specific ancestry*
[14]*Ancestry in which the allele*[13] *is enriched; Cell is colored based on the enriched ancestry: African=blue, European=yellow, Native American=red*

Table 6. Lists of pathways that show significant enrichment of genes with mapped ancestry-enriched SNPs for each admixed Latin American population.

For each pathway with significant enrichment of genes with mapped ancestry-enriched SNPs in at least one of the four populations the overlapping genes and FDR q-values are given.

Pathway	Description	Colombia		Mexico		Peru		Puerto Rico	
		Genes[1]	*q*-value[2]	Genes[1]	*q*-value[2]	Genes[1]	*q*-value[2]	Genes[1]	*q*-value[2]
KEGG Graft Versus Host Disease	Graft-versus-host disease	HLA-DOA, HLA-C, HLA-B, IL1B, HLA-DQA1, HLA-DQA2, HLA-DQB1, IL1A, FAS, HLA-DRB1	8.97E-12	HLA-DOA, HLA-C, IL1B, HLA-DPB1, HLA-DQA1, HLA-DQA2, HLA-DQB1, FAS, IL6, HLA-DRB1	8.04E-10	HLA-DOA, TNF, HLA-C, HLA-B, IL1B, HLA-DPB1, HLA-DQA1, HLA-DQA2, HLA-DQB1, IL1A, FAS, HLA-DRB1, HLA-DRA	1.49E-12	HLA-DOA, TNF, IL1B, HLA-DQA1, HLA-DQA2, HLA-DQB1, HLA-DRA	1.31E-08
KEGG Leishmania Infection	Leishmania infection	HLA-DOA, IL1A, CR1, IL10, IL1B, HLA-DQA1, HLA-DQA2, HLA-DQB1, TLR2, CYBA, HLA-DRB1	3.54E-11	HLA-DOA, IL10, IL1B, HLA-DPB1, HLA-DQA1, HLA-DQA2, HLA-DQB1, TLR4, TLR2, FOS, HLA-DRB1	3.98E-09	HLA-DOA, PRKCB, TNF, IL1A, CR1, FCGR2A, IL10, TGFB1, IL1B, HLA-DPB1, HLA-DQA1, HLA-DQA2, HLA-DQB1, TLR4, IL4, TLR2, FOS, HLA-DRB1, HLA-DRA	1.34E-16	HLA-DOA, TNF, CR1, IL1B, HLA-DQA1, HLA-DQA2, HLA-DQB1, HLA-DRA	1.31E-08

Table 6 (continued).

KEGG Type I Diabetes Mellitus	Type I diabetes mellitus	HLA-DOA, HLA-C, HLA-B, IL1B, HLA-DQA1, HLA-DQA2, HLA-DQB1, IL1A, FAS, HLA-DRB1	9.99E-12	HLA-DOA, HLA-C, IL1B, HLA-DPB1, HLA-DQA1, HLA-DQA2, HLA-DQB1, FAS, HLA-DRB1	1.22E-08	HLA-DOA, TNF, HLA-C, HLA-B, IL1B, HLA-DPB1, HLA-DQA1, HLA-DQA2, HLA-DQB1, IL1A, FAS, HLA-DRB1, HLA-DRA	2.67E-12	HLA-DOA, TNF, IL1B, HLA-DQA1, HLA-DQA2, HLA-DQB1, HLA-DRA	1.31E-08
KEGG Allograft Rejection	Allograft rejection	HLA-DOA, HLA-C, HLA-B, HLA-DQA1, HLA-DQA2, HLA-DQB1, FAS, IL10, HLA-DRB1	5.16E-11	HLA-DOA, HLA-C, HLA-DPB1, HLA-DQA1, HLA-DQA2, HLA-DQB1, FAS, IL10, HLA-DRB1	3.69E-09	HLA-DOA, TNF, HLA-C, HLA-B, HLA-DPB1, HLA-DQA1, HLA-DQA2, HLA-DQB1, FAS, IL4, IL10, HLA-DRB1, HLA-DRA	3.75E-13	HLA-DOA, TNF, HLA-DQA1, HLA-DQA2, HLA-DQB1, HLA-DRA	1.73E-07
KEGG Asthma	Asthma	HLA-DOA, HLA-DQA1, HLA-DQA2, HLA-DQB1, IL10, HLA-DRB1	5.05E-07	HLA-DOA, HLA-DPB1, HLA-DQA1, HLA-DQA2, HLA-DQB1, IL10, HLA-DRB1	2.36E-07	HLA-DOA, TNF, HLA-DPB1, HLA-DQA1, HLA-DQA2, HLA-DQB1, IL4, IL10, HLA-DRB1, HLA-DRA	3.68E-10	HLA-DOA, TNF, HLA-DQA1, HLA-DQA2, HLA-DQB1, HLA-DRA	5.31E-08

Table 6 (continued).

KEGG Metabolism Of Xenobiotics By Cytochrome P450	Metabolism of xenobiotics by cytochrome P450	EPHX1, ADH1B, CYP2C9, CYP2C19, CYP2C8, CYP2B6, UGT1A4, UGT1A1, UGT1A6, CYP1A2, UGT1A9	3.20E-11	EPHX1, CYP3A4, GSTP1, ADH1B, CYP2C9, CYP2C19, CYP2C8, CYP2B6, UGT1A1, UGT1A8, UGT1A9	3.69E-09	CYP3A5, CYP3A4, GSTP1, GSTT1, ADH1B, CYP2C9, CYP2C19, CYP2C8, CYP2B6, UGT1A1, UGT1A8, UGT1A6, CYP1A2, UGT1A9, UGT2B10, UGT2B7	3.75E-13	EPHX1, CYP3A4, ADH1B, CYP2C8, CYP2B6, UGT2B10	5.84E-06
KEGG Viral Myocarditis	Viral myocarditis	HLA-DOA, ICAM1, HLA-C, HLA-B, CCND1, HLA-DQA1, HLA-DQA2, HLA-DQB1, HLA-DRB1	1.80E-08	HLA-DOA, ICAM1, HLA-C, CCND1, HLA-DPB1, HLA-DQA1, HLA-DQA2, HLA-DQB1, MYH7, HLA-DRB1	5.22E-08	HLA-DOA, ICAM1, HLA-C, HLA-B, CCND1, HLA-DPB1, HLA-DQA1, HLA-DQA2, HLA-DQB1, MYH6, HLA-DRB1, HLA-DRA	1.99E-08	HLA-DOA, ICAM1, HLA-DQA1, HLA-DQA2, HLA-DQB1, HLA-DRA	6.77E-06

Table 6 (continued).

KEGG Neuroactive Ligand Receptor Interaction	Neuroactive ligand-receptor interaction	PTGER2, THRB, DRD2, BDKRB2, TSHR, MC1R, LEPR, MC4R, HTR2A, GIPR, OPRM1, GRIN3A, HTR1A, CHRNA4, LTB4R, TBXA2R	3.30E-10	PTGER4, PTGER3, ADRB2, ADRB1, ADRA2A, THRB, DRD2, BDKRB2, AGTR1, TSHR, GRID1, MC1R, LEPR, ADORA1, HTR4, HTR2A, GIPR, OPRM1, CRHR2, HTR1B, HTR1A, HRH3, CHRNA4, LTB4R	1.07E-13	PTGER4, PTGER3, HRH4, ADRB1, ADRA2A, THRB, DRD1, DRD2, DRD3, BDKRB2, AGTR1, TSHR, MC1R, LEPR, MC4R, ADORA1, HTR4, HTR2A, GIPR, OPRM1, OPRD1, GABRA1, GRIN3A, CRHR2, CRHR1, HTR1B, HTR1A, HRH3, CHRNA5, CHRNA4, GRIN1, LTB4R	1.10E-16	ADRB1, LEPR, OPRM1, OPRK1, F2, CHRNA4, GRIN2A, CHRNA3, LTB4R	1.28E-05
KEGG Hematopoietic Cell Lineage	Hematopoietic cell lineage	CD36, KITLG, IL1A, CR1, CD44, IL1B, HLA-DRB1	1.57E-05	CD36, KITLG, CD9, CD44, IL1B, IL6R, ITGA2, IL4R, IL6, HLA-DRB1	2.51E-07	GP1BA, TNF, CD36, KITLG, THPO, IL1A, CR1, CD9, CD44, IL1B, IL6R, ITGA2, IL4, IL4R, HLA-DRB1, HLA-DRA	1.28E-11	TNF, CD36, ITGA2B, CR1, CD9, CD44, IL1B, IL6R, ITGA2, HLA-DRA	1.69E-10

Table 6 (continued).

KEGG Intestinal Immune Network For IgA Production	Intestinal immune network for IgA production	HLA-DOA, HLA-DQA1, HLA-DQA2, HLA-DQB1, IL10, HLA-DRB1	7.96E-06	HLA-DOA, IL15, HLA-DPB1, HLA-DQA1, HLA-DQA2, HLA-DQB1, IL10, IL6, HLA-DRB1	2.15E-08	HLA-DOA, IL15, TGFB1, HLA-DPB1, HLA-DQA1, HLA-DQA2, HLA-DQB1, IL4, IL10, HLA-DRB1, HLA-DRA	2.63E-09	HLA-DOA, HLA-DQA1, HLA-DQA2, HLA-DQB1, HLA-DRA	1.75E-05
KEGG Autoimmune Thyroid Disease	Autoimmune thyroid disease	HLA-DOA, TSHR, HLA-C, HLA-B, IL10, HLA-DQA1, HLA-DQA2, HLA-DQB1, FAS, HLA-DRB1	3.75E-11	HLA-DOA, TSHR, HLA-C, IL10, HLA-DPB1, HLA-DQA1, HLA-DQA2, HLA-DQB1, FAS, HLA-DRB1	3.69E-09	HLA-DOA, TSHR, HLA-C, HLA-B, IL10, HLA-DPB1, HLA-DQA1, HLA-DQA2, HLA-DQB1, FAS, IL4, HLA-DRB1, HLA-DRA	3.00E-11	HLA-DOA, HLA-DQA1, HLA-DQA2, HLA-DQB1, HLA-DRA	2.69E-05
KEGG Arachidonic Acid Metabolism	Arachidonic acid metabolism	CYP2C9, CYP2C19, CYP2C8, CBR1, CYP2B6, PTGES	1.88E-05	GGT1, PLA2G3, PLA2G4A, PTGS1, CYP2C9, CYP2C19, CYP2C8, CYP2B6, PTGES	9.07E-08	PLA2G3, CYP4F2, PLA2G4A, PTGS1, CYP2C9, CYP2C19, CYP2C8, CBR1, CYP2B6, CBR3, LTA4H, PTGES	1.44E-09	PLA2G3, CYP2C8, CBR1, CYP2B6, PTGES, TBXAS1	2.10E-06

Table 6 (continued).

KEGG Retinol Metabolism	Retinol metabolism	ADH1B, CYP2C9, CYP2C19, CYP2C8, CYP2B6, UGT1A4, UGT1A1, UGT1A6, CYP1A2, UGT1A9	2.07E-10	CYP3A4, ADH1B, CYP26A1, CYP2C9, CYP2C19, CYP2C8, CYP2B6, UGT1A1, UGT1A8, UGT1A9	1.72E-08	CYP3A5, CYP3A4, ADH1B, CYP26A1, CYP2C9, CYP2C19, CYP2C8, CYP2B6, UGT1A1, UGT1A8, UGT1A6, CYP2A6, CYP1A2, UGT1A9, UGT2B10, UGT2B7	1.21E-13	CYP3A4, ADH1B, CYP2C8, CYP2B6, UGT2B10	6.48E-05
KEGG Drug Metabolism Cytochrome P450	Drug metabolism - cytochrome P450	CYP2D6, ADH1B, FMO1, CYP2C9, CYP2C19, CYP2C8, CYP2B6, UGT1A4, UGT1A1, UGT1A6, CYP1A2, UGT1A9	5.01E-12	CYP3A4, GSTP1, CYP2D6, ADH1B, FMO1, FMO3, CYP2C9, CYP2C19, CYP2C8, CYP2B6, UGT1A1, UGT1A8, UGT1A9	3.63E-11	CYP3A5, CYP3A4, GSTP1, GSTT1, CYP2D6, ADH1B, FMO1, FMO3, CYP2C9, CYP2C19, CYP2C8, CYP2B6, UGT1A1, UGT1A8, UGT1A6, CYP2A6, CYP1A2, UGT1A9, UGT2B10, UGT2B7	2.13E-17	CYP3A4, ADH1B, CYP2C8, CYP2B6, UGT2B10	1.03E-04

Table 6 (continued).

KEGG Cell Adhesion Molecules CAMS	Cell adhesion molecules (CAMs)	CD276, HLA-DQA1, HLA-DQA2, HLA-DQB1, HLA-DRB1, PVRL2, HLA-DOA, ICAM1, HLA-C, HLA-B	2.32E-07	ITGA9, CD276, HLA-DPB1, HLA-DQA1, HLA-DQA2, HLA-DQB1, HLA-DRB1, HLA-DOA, ICAM1, HLA-C	9.83E-06	CDH5, SELP, HLA-DPB1, HLA-DQA1, HLA-DQA2, HLA-DQB1, HLA-DRB1, PVRL2, HLA-DRA, HLA-DOA, ICAM1, HLA-C, HLA-B, NEGR1	2.77E-07	HLA-DQA1, HLA-DQA2, HLA-DQB1, HLA-DRA, HLA-DOA, ICAM1	1.27E-04
KEGG Cytokine Cytokine Receptor Interaction	Cytokine-cytokine receptor interaction	KITLG, LEPR, CXCL13, IL1A, ACVR2B, TNFRSF1B, TNFRSF1A, IL10, IL1B, BMP2, IL28B, IL29, FAS, IL28A	2.54E-08	CTF1, IL23R, INHBC, KITLG, LEPR, IL18, IL15, CXCL13, TNFRSF1B, IL10, TNFRSF10A, IL1B, IL6R, IL28B, IL29, FAS, IL28A, IL4R, IL6	1.58E-09	CCL2, CTF1, TGFB1, CXCL5, INHBC, INHBA, KITLG, LEPR, IL18, IL15, CXCL13, IL1A, ACVR2B, TNF, IL8, TNFRSF1B, TNFRSF1A, IL10, TNFRSF10A, IL1B, IL6R, IL28B, IL29, FAS, IL28A, IL4, IL4R	1.21E-13	LEPR, TNF, CXCR2, CXCR1, TNFRSF1A, IL1B, IL6R	5.26E-04

Table 6 (continued).

KEGG Antigen Processing And Presentation	Antigen processing and presentation	HLA-DOA, HLA-C, HLA-B, HLA-DQA1, HLA-DQA2, HLA-DQB1, HLA-DRB1	1.62E-05	HLA-DOA, HLA-C, HLA-DPB1, HLA-DQA1, HLA-DQA2, HLA-DQB1, HLA-DRB1	2.23E-04	HLA-DOA, HLA-C, HLA-B, HSPA5, HLA-DPB1, HLA-DQA1, HLA-DQA2, HLA-DQB1, HLA-DRB1, HLA-DRA	9.25E-06	HLA-DOA, HLA-DQA1, HLA-DQA2, HLA-DQB1, HLA-DRA	2.25E-04
KEGG Mapk Signaling Pathway	MAPK signaling pathway	MAP3K5, MAP3K1, BDNF, FGF2, IL1A, TAOK1, CACNA1C, TNFRSF1A, IL1B, FAS	6.19E-05	MEF2C, MAP3K1, AKT1, BDNF, MAPK10, PLA2G4A, FOS, NGF, TAOK1, PLA2G3, CACNA1C, IL1B, FAS	2.53E-05	MEF2C, MAP3K5, MAP3K1, MAP3K3, AKT1, TGFB1, BDNF, MAPK10, PLA2G4A, CACNG2, MOS, CACNA1I, FOS, IL1A, NGF, TAOK1, PRKCB, TNF, PLA2G3, CACNA1D, CACNA1C, TNFRSF1A, IL1B, FAS	1.37E-10	AKT1, BDNF, TNF, PLA2G3, CACNA1C, TNFRSF1A, IL1B	5.26E-04

Table 6 (continued).

KEGG Adipocytokine Signaling Pathway	Adipocytokine signaling pathway	CD36, RXRG, TNFRSF1B, TNFRSF1A, LEPR	4.43E-04	PPARGC1A, CD36, MAPK10, RXRB, RXRG, TNFRSF1B, AKT1, LEPR	3.97E-06	PTPN11, TNF, CD36, MAPK10, RXRB, RXRG, TNFRSF1B, AKT1, TNFRSF1A, LEPR, PPARA	9.19E-08	TNF, CD36, AKT1, TNFRSF1A, LEPR	7.64E-05
KEGG Alzheimers Disease	Alzheimer's disease	ATP2A1, APOE, CACNA1C, TNFRSF1A, GSK3B, IL1B, FAS	5.59E-04	ATP2A1, LRP1, APOE, ATF6, BAD, CACNA1C, NDUFAB1, GSK3B, IL1B, FAS	6.07E-05	ATP2A1, LRP1, COX4I1, APOE, LPL, ATF6, TNF, BAD, GRIN1, CACNA1D, CACNA1C, TNFRSF1A, GSK3B, IL1B, FAS	6.98E-07	APOE, TNF, GRIN2A, CACNA1C, TNFRSF1A, IL1B	3.94E-04
KEGG Linoleic Acid Metabolism	Linoleic acid metabolism	CYP1A2, CYP2C9, CYP2C19, CYP2C8	2.47E-04	CYP3A4, PLA2G3, PLA2G4A, CYP2C9, CYP2C19, CYP2C8	4.31E-06	CYP3A5, CYP3A4, PLA2G3, CYP1A2, PLA2G4A, CYP2C9, CYP2C19, CYP2C8	1.31E-07	CYP3A4, PLA2G3, CYP2C8	1.82E-03

Table 6 (continued).

KEGG Dilated Cardiomyopathy	Dilated cardiomyopathy	CACNA1C, ADCY4, TNNT2, ADCY3, GNAS	1.55E-03	ITGA9, RYR2, ADRB1, CACNA1C, ADCY4, TNNT2, MYL2, ADCY3, ITGA2, GNAS, MYH7	3.84E-08	ADCY8, TNF, IGF1, RYR2, ADRB1, CACNA1D, CACNA1C, ADCY4, CACNG2, TNNT2, MYL2, TGFB1, ADCY3, ITGA2, GNAS, MYH6	2.42E-11	ADCY8, TNF, ADRB1, TPM1, ITGA2B, CACNA1C, ADCY4, ITGA2	5.88E-08
KEGG Calcium Signaling Pathway	Calcium signaling pathway	ATP2A1, ADCY3, BDKRB2, GNAS, CACNA1C, ADCY4, HTR2A, TBXA2R	1.30E-04	PTGER3, ATP2A1, ADRB2, ADRB1, ADCY3, BDKRB2, GNAS, AGTR1, RYR2, CACNA1C, ADCY4, HTR4, HTR2A	3.60E-07	PTGER3, ATP2A1, ADRB1, DRD1, ADCY3, BDKRB2, GNAS, AGTR1, ADCY8, PRKCB, RYR2, GRIN1, NOS3, CACNA1D, CACNA1C, ADCY4, HTR4, HTR2A, CACNA1I	9.77E-10	ADRB1, ADCY8, GRIN2A, CACNA1C, ADCY4	4.28E-03

Table 6 (continued).

KEGG Apoptosis	Apoptosis	IL1A, TNFRSF1A, IL1B, FAS, EXOG	1.31E-03	BAD, AKT1, TNFRSF10A, NGF, IL1B, IRAK3, FAS	2.12E-04	TNF, BAD, IL1A, AKT1, TNFRSF1A, TNFRSF10A, NGF, IL1B, IRAK3, FAS, EXOG	1.16E-06	TNF, AKT1, TNFRSF1A, IL1B	2.90E-03
KEGG Systemic Lupus Erythematosus	Systemic lupus erythematosus	HLA-DQA1, HLA-DQA2, HLA-DQB1, HLA-DRB1, HLA-DOA, IL10	1.35E-03	HLA-DPB1, HLA-DQA1, HLA-DQA2, HLA-DQB1, HLA-DRB1, HLA-DOA, IL10	2.53E-03	FCGR2A, HLA-DPB1, HLA-DQA1, HLA-DQA2, HLA-DQB1, HLA-DRB1, HLA-DRA, HLA-DOA, TNF, IL10	3.07E-04	HLA-DQA1, HLA-DQA2, HLA-DQB1, HLA-DRA, HLA-DOA, TNF, GRIN2A	1.57E-05
KEGG Drug Metabolism Other Enzymes	Drug metabolism - other enzymes	UGT1A4, UGT1A1, UGT1A6, DPYD, CES1, UGT1A9	1.09E-05	CYP3A4, CDA, TPMT, UGT1A1, UGT1A8, TYMP, DPYD, UMPS, UGT1A9, NAT2	3.69E-09	XDH, CYP3A5, CYP3A4, CDA, TPMT, UGT1A1, UGT1A8, ITPA, UGT1A6, CYP2A6, DPYD, UMPS, CES1, UGT1A9, NAT2, UGT2B10, UGT2B7	1.33E-16	CYP3A4, TPMT, UGT2B10	6.97E-03

Table 6 (continued).

KEGG GnRH Signaling Pathway	GnRH signaling pathway	MAP3K1, ADCY3, GNAS, CACNA1C, ADCY4	2.25E-03	MAP3K1, ADCY3, GNAS, PLA2G3, MAPK10, PLA2G4A, CACNA1C, ADCY4	6.11E-05	MAP3K1, MAP3K3, ADCY3, GNAS, ADCY8, PRKCB, PLA2G3, MAPK10, CACNA1D, PLA2G4A, CACNA1C, ADCY4	6.11E-07	ADCY8, PLA2G3, CACNA1C, ADCY4	4.56E-03
KEGG Steroid Hormone Biosynthesis	Steroid hormone biosynthesis	CYP7A1, CYP19A1, UGT1A4, UGT1A1, UGT1A6, UGT1A9	1.56E-05	COMT, CYP3A4, CYP7A1, CYP19A1, CYP11B2, UGT1A1, UGT1A8, HSD17B8, UGT1A9	6.25E-08	CYP3A5, COMT, CYP3A4, CYP7A1, CYP19A1, CYP11B2, UGT1A1, UGT1A8, UGT1A6, HSD17B8, CYP7B1, UGT1A9, HSD17B12, UGT2B10, UGT2B7	1.89E-13	CYP3A4, CYP7A1, UGT2B10	8.06E-03

Table 6 (continued).

KEGG Vascular Smooth Muscle Contraction	Vascular smooth muscle contraction	ADCY3, GNAS, KCNMA1, CACNA1C, ADCY4	3.72E-03	ADCY3, GNAS, AGTR1, PLA2G3, PLA2G4A, CACNA1C, ADCY4	8.32E-04	ADCY3, GNAS, AGTR1, ADCY8, PRKCB, KCNMA1, PLA2G3, PPP1R12A, CACNA1D, PLA2G4A, CACNA1C, ADCY4	2.15E-06	ADCY8, PLA2G3, CACNA1C, ADCY4	6.51E-03
KEGG PPAR Signaling Pathway	PPAR signaling pathway	CD36, FADS2, CYP7A1, RXRG, FABP1, PPARG, PPARD	4.02E-06	ACADM, CD36, FADS2, CYP7A1, RXRB, RXRG, ACOX3, FABP1, PPARD	3.60E-07	ACADL, ACADM, CD36, FADS2, CYP7A1, APOA5, RXRB, RXRG, ACOX3, LPL, FABP1, PPARA, PPARD	9.08E-10	CD36, FADS2, CYP7A1	1.33E-02
KEGG Jak STAT Signaling Pathway	Jak-STAT signaling pathway	CCND1, LEPR, IL10, CBL, SPRY1, STAM2, IL28B, IL29, IL28A	7.96E-06	CTF1, AKT1, IL23R, CCND1, LEPR, IL10, IL15, IL6R, STAM2, IL28B, IL29, IL28A, IL4R, IL6	1.23E-08	CTF1, AKT1, CCND1, PTPN11, LEPR, IL10, CBL, IL15, IL6R, STAM2, IL28B, IL29, IL28A, IL4, IL4R	4.08E-08	AKT1, LEPR, IL6R, STAM2	1.40E-02
BioCarta IL5 Pathway	IL 5 Signaling Pathway	IL1B, HLA-DRB1	8.24E-03	IL1B, IL6, HLA-DRB1	7.10E-04	IL4, IL1B, HLA-DRB1, HLA-DRA	6.40E-05	IL1B, HLA-DRA	5.53E-03

Table 6 (continued).

BioCarta PML Pathway	Regulation of transcriptional activity by PML	TNFRSF1B, TNFRSF1A, FAS	1.01E-03	PRAM1, TNFRSF1B, FAS	3.05E-03	PRAM1, TNF, TNFRSF1B, TNFRSF1A, FAS	3.01E-05	TNF, TNFRSF1A	1.25E-02
KEGG Complement And Coagulation Cascades	Complement and coagulation cascades	CR1, F12, F13A1, BDKRB2, CFH	4.87E-04	VWF, F13A1, FGG, BDKRB2, FGA	3.04E-03	VWF, F3, CR1, F12, F13A1, FGG, BDKRB2, CFH, FGA	9.13E-06	F2, VWF, CR1	1.33E-02
KEGG Gap Junction	Gap junction	ADCY4, DRD2, GJA1, HTR2A, ADCY3, GNAS	1.71E-04	ADRB1, ADCY4, DRD2, GJA1, HTR2A, ADCY3, GNAS	2.35E-04	ADCY8, PRKCB, ADRB1, ADCY4, DRD1, DRD2, GJA1, HTR2A, ADCY3, GNAS	1.01E-05	ADCY8, ADRB1, ADCY4	2.13E-02
BioCarta NFkB Pathway	NF-kB Signaling Pathway	MAP3K1, IL1A, TNFRSF1B, TNFRSF1A	1.16E-04	MAP3K1, TNFRSF1B, TLR4	5.91E-03	MAP3K1, TNF, IL1A, TNFRSF1B, TNFRSF1A, TLR4	8.74E-06	TNF, TNFRSF1A	1.79E-02
BioCarta Inflam Pathway	Cytokines and Inflammatory Response	IL1A, IL10, HLA-DRB1	3.95E-03	IL15, IL10, IL6, HLA-DRB1	1.07E-03	TNF, IL15, TGFB1, IL8, IL1A, IL4, IL10, HLA-DRB1, HLA-DRA	6.22E-09	TNF, HLA-DRA	2.54E-02
KEGG T Cell Receptor Signaling Pathway	T cell receptor signaling pathway	NFAT5, IL10, GSK3B, CBL, CDK4	2.98E-03	AKT1, NFAT5, NCK2, IL10, GSK3B, CDK4, FOS	6.14E-04	AKT1, NFAT5, NCK2, TNF, IL10, GSK3B, CBL, CDK4, IL4, FOS	4.35E-05	AKT1, TNF, CDK4	3.24E-02

Table 6 (continued).

BioCarta IL1R Pathway	Signal transduction through IL1R	IL1B, MAP3K1, IL1A	5.53E-03	IL1B, MAP3K1, IRAK3, IL6	1.71E-03	TNF, TGFB1, IL1B, MAP3K1, IRAK3, IL1A	6.15E-05	TNF, IL1B	3.13E-02
KEGG Porphyrin And Chlorophyll Metabolism	Porphyrin and chlorophyll metabolism	UGT1A6, UGT1A9, FTH1, UGT1A4, UGT1A1	6.13E-05	UGT1A8, UGT1A9, FTH1, UGT1A1	3.48E-03	UGT1A8, UGT1A6, UGT1A9, FTH1, UGT2B10, UGT2B7, UGT1A1	2.07E-05	FTH1, UGT2B10	4.58E-02
KEGG Pathways In Cancer	Pathways in cancer	KITLG, FGF2, CBL, CCND1, RXRG, GSK3B, CDKN2A, CDK4, BMP2, PPARG, PPARD, CDK6, FAS, HDAC1	3.00E-07	GSTP1, AKT1, KITLG, BAD, MAPK10, FOS, CCND1, RXRB, RXRG, GSK3B, CDKN2A, ITGA2, CDK4, PPARD, CDK6, FAS, HDAC1, IL6	8.30E-08	IGF1, GSTP1, FZD8, AKT1, TGFB1, KITLG, BAD, MAPK10, CBL, CTNNA3, FOS, CCND1, PRKCB, IL8, RXRB, RXRG, GSK3B, CDKN2A, ITGA2, CDK4, PPARD, CDK6, FAS, HDAC1	6.72E-09	NA	NA
BioCarta G1 Pathway	Cell Cycle: G1/S Check Point	GSK3B, CDKN2A, CCND1, CDK4, CDK6, HDAC1	3.46E-07	GSK3B, CDKN2A, CCND1, CDK4, CDK6, HDAC1	3.74E-06	GSK3B, CDKN2A, TGFB1, CCND1, CDK4, CDK6, DHFR, HDAC1	1.03E-07	NA	NA

Table 6 (continued).

KEGG Chronic Myeloid Leukemia	Chronic myeloid leukemia	CDKN2A, CBL, CCND1, CDK4, CDK6, HDAC1	6.29E-05	BAD, AKT1, CDKN2A, CCND1, CDK4, CDK6, HDAC1	6.47E-05	PTPN11, BAD, AKT1, GAB2, CDKN2A, CBL, TGFB1, CCND1, CDK4, CDK6, HDAC1	2.02E-07	NA	NA
KEGG Hypertrophic Cardiomyopathy HCM	Hypertrophic cardiomyopathy (HCM)	NA	NA	ITGA9, RYR2, CACNA1C, TNNT2, MYL2, ITGA2, MYH7, IL6	2.04E-05	TNF, IGF1, RYR2, CACNA1D, CACNA1C, CACNG2, TNNT2, MYL2, TGFB1, ITGA2, MYH6, ACE	1.03E-07	TNF, TPM1, ITGA2B, CACNA1C, ITGA2	1.88E-04
KEGG Non Small Cell Lung Cancer	Non-small cell lung cancer	RXRG, CDKN2A, CCND1, CDK4, CDK6	1.76E-04	BAD, RXRB, RXRG, AKT1, CDKN2A, CCND1, CDK4, CDK6	7.66E-07	PRKCB, BAD, RXRB, RXRG, AKT1, CDKN2A, CCND1, CDK4, CDK6	1.30E-06	NA	NA
BioCarta BLymphocyte Pathway	B Lymphocyte Cell Surface Molecules	CR1, HLA-DRB1, ICAM1	3.16E-04	NA	NA	CR1, HLA-DRB1, HLA-DRA, ICAM1	9.64E-05	CR1, HLA-DRA, ICAM1	1.27E-04
KEGG ECM Receptor Interaction	ECM-receptor interaction	NA	NA	VWF, ITGA9, CD36, SV2C, SV2A, CD44, ITGA2, COL11A2	1.92E-05	GP1BA, VWF, CD36, SV2C, SV2A, CD44, ITGA2, COL11A2	2.35E-04	VWF, CD36, ITGA2B, CD44, ITGA2	1.87E-04

Table 6 (continued).

BioCarta WNT Pathway	WNT Signaling Pathway	GSK3B, CCND1, PPARD, HDAC1	1.71E-04	GSK3B, WIF1, CCND1, PPARD, HDAC1	4.97E-05	GSK3B, WIF1, CCND1, PPARD, HDAC1	2.24E-04	NA	NA
KEGG Ascorbate And Aldarate Metabolism	Ascorbate and aldarate metabolism	UGT1A6, UGT1A9, UGT1A4, UGT1A1	1.54E-04	UGT1A8, ALDH2, UGT1A9, UGT1A1	6.44E-04	UGT1A8, UGT1A6, ALDH2, UGT1A9, UGT2B10, UGT2B7, UGT1A1	8.13E-07	NA	NA
BioCarta ASBcell Pathway	Antigen Dependent B Cell Activation	FAS, IL10, HLA-DRB1	4.09E-04	FAS, IL10, HLA-DRB1	1.16E-03	FAS, IL4, IL10, HLA-DRB1, HLA-DRA	5.28E-06	NA	NA
KEGG Melanoma	Melanoma	FGF2, CDKN2A, CCND1, CDK4, CDK6	5.46E-04	BAD, AKT1, CDKN2A, CCND1, CDK4, CDK6	4.75E-04	IGF1, BAD, AKT1, CDKN2A, CCND1, CDK4, CDK6	4.93E-04	NA	NA
KEGG Melanogenesis	Melanogenesis	EDN1, ADCY3, GNAS, KITLG, MC1R, ADCY4, GSK3B	3.75E-05	ADCY3, GNAS, KITLG, MC1R, ADCY4, GSK3B	2.69E-03	FZD8, EDN1, ADCY3, GNAS, ADCY8, PRKCB, KITLG, MC1R, ADCY4, GSK3B	2.83E-05	NA	NA
BioCarta Gsk3 Pathway	Inactivation of Gsk3 by AKT causes accumulation of b-catenin in Alveolar Macrophages	GSK3B, GJA1, CCND1	3.40E-03	GSK3B, GJA1, CCND1, AKT1, TLR4	5.72E-05	GSK3B, GJA1, CCND1, AKT1, TLR4	2.60E-04	NA	NA

Table 6 (continued).

KEGG P53 Signaling Pathway	p53 signaling pathway	CDKN2A, CCND1, CDK4, FAS, CDK6	4.87E-04	CDKN2A, CCND1, CDK4, FAS, CDK6	3.04E-03	IGF1, RRM2B, SERPINB5, CDKN2A, CCND1, CDK4, FAS, CDK6	6.40E-05	NA	NA
KEGG Pancreatic Cancer	Pancreatic cancer	CDKN2A, CCND1, CDK4, CDK6	4.62E-03	BAD, MAPK10, AKT1, CDKN2A, CCND1, CDK4, CDK6	5.44E-05	BAD, MAPK10, AKT1, CDKN2A, TGFB1, CCND1, CDK4, CDK6	6.96E-05	NA	NA
KEGG Neurotrophin Signaling Pathway	Neurotrophin signaling pathway	MAP3K5, MAP3K1, SH2B1, BDNF, GSK3B	5.28E-03	MAP3K1, AKT1, SH2B1, NGF, BDNF, BAD, MAPK10, GSK3B, IRAK3	4.33E-05	MAP3K5, MAP3K1, MAP3K3, AKT1, SH2B1, NGF, BDNF, SORT1, PTPN11, BAD, MAPK10, GSK3B, IRAK3	8.56E-07	NA	NA
BioCarta Keratinocyte Pathway	Keratinocyte Differentiation	MAP3K5, MAP3K1, TNFRSF1B, TNFRSF1A, FAS	9.82E-05	MAP3K1, TNFRSF1B, FAS, FOS	4.79E-03	PRKCB, TNF, MAP3K5, MAP3K1, TNFRSF1B, TNFRSF1A, FAS, FOS	4.04E-06	NA	NA

Table 6 (continued).

KEGG Amyotrophic Lateral Sclerosis ALS	Amyotrophic lateral sclerosis (ALS)	MAP3K5, TNFRSF1B, TNFRSF1A, TOMM40	1.88E-03	NA	NA	TNF, BAD, MAP3K5, GRIN1, SOD1, TNFRSF1B, TNFRSF1A, TOMM40	1.10E-05	TNF, GRIN2A, TNFRSF1A	7.61E-03
KEGG Glioma	Glioma	CDKN2A, CCND1, CDK4, CDK6	3.72E-03	AKT1, CDKN2A, CCND1, CDK4, CDK6	2.49E-03	PRKCB, IGF1, AKT1, CDKN2A, CCND1, CDK4, CDK6	3.04E-04	NA	NA
BioCarta IL2RB Pathway	IL-2 Receptor Beta Chain in T cell Activation	CBL, SYK, FAS	8.13E-03	SYK, BAD, AKT1, FAS, FOS	2.55E-04	CBL, SYK, BAD, AKT1, FAS, FOS	1.28E-04	NA	NA
KEGG Small Cell Lung Cancer	Small cell lung cancer	RXRG, CCND1, CDK4, CDK6	8.24E-03	RXRB, RXRG, AKT1, CCND1, ITGA2, CDK4, CDK6	1.60E-04	NA	NA	ITGA2B, AKT1, ITGA2, CDK4	2.60E-03
BioCarta Toll Pathway	Toll-Like Receptor Pathway	MAP3K1, TLR3, TLR2	7.63E-03	MAP3K1, TLR4, FOS, TLR2	2.55E-03	PPARA, MAP3K1, TLR4, FOS, TLR2	9.64E-04	NA	NA
KEGG Alpha Linolenic Acid Metabolism	alpha-Linolenic acid metabolism	NA	NA	PLA2G4A, FADS2, ACOX3, PLA2G3	2.48E-04	PLA2G4A, FADS2, ACOX3, PLA2G3	7.43E-04	FADS2, PLA2G3	1.40E-02

Table 6 (continued).

KEGG Natural Killer Cell Mediated Cytotoxicity	Natural killer cell mediated cytotoxicity	NFAT5, ICAM1, HLA-C, HLA-B, SYK, FAS	1.25E-03	NFAT5, ICAM1, HLA-C, TNFRSF10A, SYK, FAS	8.30E-03	NFAT5, PTPN11, PRKCB, TNF, ICAM1, HLA-C, HLA-B, TNFRSF10A, SYK, FAS	2.67E-04	NA	NA
KEGG Pentose And Glucuronate Interconversions	Pentose and glucuronate interconversions	UGT1A6, UGT1A9, XYLB, UGT1A4, UGT1A1	1.11E-05	UGT1A8, UGT1A9, UGT1A1	9.39E-03	UGT1A8, UGT1A6, UGT1A9, XYLB, UGT2B10, UGT2B7, UGT1A1	1.73E-06	NA	NA
KEGG Arrhythmogenic Right Ventricular Cardiomyopathy ARVC	Arrhythmogenic right ventricular cardiomyopathy (ARVC)	NA	NA	ITGA9, RYR2, CACNA1C, GJA1, ITGA2	4.15E-03	RYR2, CACNA1D, CACNA1C, CACNG2, GJA1, ITGA2, CTNNA3	7.10E-04	ITGA2B, CACNA1C, ITGA2	1.59E-02
KEGG Fc Epsilon Ri Signaling Pathway	Fc epsilon RI signaling pathway	NA	NA	PLA2G3, MAPK10, PLA2G4A, AKT1, SYK	4.79E-03	PRKCB, TNF, PLA2G3, MAPK10, PLA2G4A, AKT1, GAB2, SYK, IL4, LYN	3.35E-06	TNF, PLA2G3, AKT1	1.70E-02
BioCarta Granulocytes Pathway	Adhesion and Diapedesis of Granulocytes	IL1A, ICAM1	1.44E-02	NA	NA	TNF, IL1A, ICAM1, SELP, IL8	1.15E-05	TNF, ICAM1	9.24E-03
BioCarta RACCYCD Pathway	Influence of Ras and Rho proteins on G1 to S Transition	CCND1, CDK4, CDK6	3.09E-03	CCND1, CDK4, AKT1, CDK6	7.29E-04	NA	NA	CDK4, AKT1	2.13E-02

Table 6 (continued).

KEGG Focal Adhesion	Focal adhesion	NA	NA	AKT1, MYL2, COL11A2, VWF, BAD, MAPK10, ITGA9, PARVB, CCND1, BCAR1, GSK3B, ITGA2	8.84E-06	IGF1, AKT1, MYL2, COL11A2, VWF, PPP1R12A, BAD, MAPK10, PARVB, CCND1, BCAR1, PRKCB, GSK3B, ITGA2	2.29E-05	AKT1, VWF, ITGA2B, ITGA2	2.76E-02
KEGG Toll Like Receptor Signaling Pathway	Toll-like receptor signaling pathway	NA	NA	AKT1, MAPK10, IL1B, TLR4, FOS, TLR2, IL6	4.61E-04	AKT1, TNF, IL8, MAPK10, IL1B, TLR4, FOS, TLR2	7.44E-04	AKT1, TNF, IL1B	2.84E-02
KEGG Chemokine Signaling Pathway	Chemokine signaling pathway	ADCY3, ADCY4, GSK3B, CXCL13, GRK4	2.11E-02	NA	NA	CCL2, AKT1, GNB3, CXCL5, ADCY3, BCAR1, ADCY8, PRKCB, IL8, WASL, ADCY4, GSK3B, CXCL13, ELMO1, GRK4, LYN	5.66E-07	AKT1, ADCY8, CXCR2, CXCR1, ADCY4, ELMO1	6.72E-04
BioCarta LAIR Pathway	Cells and Molecules involved in local acute inflammatory response	IL1A, ICAM1	1.82E-02	NA	NA	TNF, IL1A, ICAM1, SELP, IL8	3.01E-05	TNF, ICAM1	1.25E-02

Table 6 (continued).

BioCarta NFAT Pathway	NFAT and Hypertrophy of the heart (Transcription in the broken heart)	FGF2, EDN1, GSK3B	1.82E-02	MEF2C, CTF1, AKT1, GSK3B	8.26E-03	MEF2C, IGF1, CTF1, AKT1, EDN1, GSK3B, AGT	1.28E-04	NA	NA
BioCarta Cytokine Pathway	Cytokine Network	NA	NA	IL18, IL15, IL10, IL6	4.29E-04	IL18, TNF, IL1A, IL15, IL4, IL10, IL8	3.39E-07	NA	NA
BioCarta Cellcycle Pathway	Cyclins and Cell Cycle Regulation	CCND1, CDKN2A, CDK4, CDK6	1.16E-04	CCND1, CDKN2A, CDK4, CDK6	4.75E-04	NA	NA	NA	NA
KEGG Colorectal Cancer	Colorectal cancer	NA	NA	BAD, MAPK10, AKT1, GSK3B, CCND1, FOS	2.48E-04	BAD, MAPK10, AKT1, GSK3B, TGFB1, CCND1, FOS	2.35E-04	NA	NA
KEGG Fatty Acid Metabolism	Fatty acid metabolism	NA	NA	ACADS, ALDH2, ACADM, ADH1B, ACOX3	4.10E-04	ACADS, ACADL, ALDH2, ACADM, ADH1B, ACOX3	2.24E-04	NA	NA
KEGG Cardiac Muscle Contraction	Cardiac muscle contraction	NA	NA	RYR2, ATP1B1, CACNA1C, TNNT2, MYL2, MYH7	8.28E-04	RYR2, ATP1B1, CACNA1D, CACNA1C, COX4I1, CACNG2, TNNT2, MYL2, MYH6	2.80E-05	NA	NA

Table 6 (continued).

BioCarta NuclearRs Pathway	Nuclear Receptors in Lipid Metabolism and Toxicity	PPARG, PPARD, CYP1A2	7.26E-04	NA	NA	ABCB1, PPARD, PPARA, CYP1A2	3.07E-04	NA	NA
BioCarta CARM Er Pathway	CARM1 and Regulation of the Estrogen Receptor	NA	NA	MEF2C, PPARGC1A, ERCC3, ESR1, CCND1, HDAC1	1.26E-05	MEF2C, ERCC3, ESR1, CCND1, HDAC1	7.60E-04	NA	NA
KEGG Purine Metabolism	Purine metabolism	NA	NA	NA	NA	XDH, ATIC, PDE4D, ADCY3, GDA, DCK, ADCY8, RRM2B, ADCY4, ITPA, RRM1, PDE8B	4.43E-05	PDE4D, GDA, ADCY8, ADCY4, RRM2	2.72E-03
BioCarta HIVNef Pathway	HIV-I Nef: negative effector of Fas and TNF	MAP3K5, MAP3K1, TNFRSF1B, TNFRSF1A, FAS	2.43E-04	NA	NA	TNF, MAP3K5, MAP3K1, TNFRSF1B, TNFRSF1A, FAS	1.02E-03	NA	NA
KEGG Starch And Sucrose Metabolism	Starch and sucrose metabolism	UGT1A4, UGT1A1, UGT1A6, UGT1A9	1.78E-03	NA	NA	UGT1A1, UGT1A8, UGT1A6, UGT1A9, UGT2B10, UGT2B7	6.19E-04	NA	NA
KEGG Endocytosis	Endocytosis	LDLR, NEDD4L, HLA-C, HLA-B, CBL, STAM2, GRK4	8.81E-04	NA	NA	NA	NA	ADRB1, LDLR, CXCR2, CXCR1, STAM2	4.56E-03

Table 6 (continued).

BioCarta FcER1 Pathway	Fc Epsilon Receptor I Signaling in Mast Cells	NA	NA	SYK, PLA2G4A, MAP3K1, FOS	3.48E-03	PRKCB, SYK, PLA2G4A, MAP3K1, FOS, LYN	1.97E-04	NA	NA
BioCarta Th1Th2 Pathway	Th1/Th2 Differentiation	NA	NA	IL18, IL4R, HLA-DRB1	4.05E-03	IL18, IL4, IL4R, HLA-DRB1, HLA-DRA	5.15E-05	NA	NA
KEGG ABC Transporters	ABC transporters	NA	NA	ABCC2, ABCC3, ABCC4, ABCG8	4.15E-03	ABCA1, ABCC2, ABCC6, ABCG2, ABCC4, ABCB1, ABCC1, ABCG8	2.95E-06	NA	NA
KEGG B Cell Receptor Signaling Pathway	B cell receptor signaling pathway	NA	NA	AKT1, NFAT5, GSK3B, SYK, FOS	4.09E-03	PRKCB, AKT1, NFAT5, GSK3B, SYK, FOS, LYN	6.68E-04	NA	NA
KEGG VEGF Signaling Pathway	VEGF signaling pathway	NA	NA	PLA2G3, BAD, PLA2G4A, AKT1, NFAT5	4.15E-03	PRKCB, PLA2G3, BAD, NOS3, PLA2G4A, AKT1, NFAT5	7.10E-04	NA	NA
BioCarta Longevity Pathway	The IGF-1 Receptor and Longevity	NA	NA	NA	NA	SOD1, SOD2, IGF1, AKT1	3.07E-04	SOD2, AKT1	1.04E-02
KEGG Cell Cycle	Cell cycle	CCND1, GSK3B, CDKN2A, CDK4, CDK6, HDAC1	8.97E-04	CCND1, GSK3B, CDKN2A, CDK4, CDK6, HDAC1	6.47E-03	NA	NA	NA	NA

Table 6 (continued).

BioCarta SODD Pathway	SODD/TNFR1 Signaling Pathway	TNFRSF1B, TNFRSF1A	8.24E-03	NA	NA	NA	NA	TNF, TNFRSF1A	5.53E-03
BioCarta Monocyte Pathway	Monocyte and its Surface Molecules	CD44, ICAM1	9.53E-03	NA	NA	NA	NA	CD44, ICAM1	6.51E-03
KEGG Bladder Cancer	Bladder cancer	CDKN2A, CCND1, CDK4	9.75E-03	CDKN2A, TYMP, CCND1, CDK4	3.76E-03	NA	NA	NA	NA
KEGG Vasopressin Regulated Water Reabsorption	Vasopressin-regulated water reabsorption	ADCY3, GNAS, AQP2	1.10E-02	ADCY3, DCTN5, STX4, GNAS	4.15E-03	NA	NA	NA	NA
BioCarta Eryth Pathway	Erythrocyte Differentiation Pathway	IL1A, KITLG	1.56E-02	NA	NA	IGF1, IL1A, TGFB1, KITLG	3.07E-04	NA	NA
BioCarta SPPA Pathway	Aspirin Blocks Signaling Pathway Involved in Platelet Activation	NA	NA	PLA2G4A, PTGS1, SYK	5.29E-03	NA	NA	F2, TBXAS1	1.70E-02
KEGG Biosynthesis Of Unsaturated Fatty Acids	Biosynthesis of unsaturated fatty acids	NA	NA	FADS1, FADS2, ACOX3	5.29E-03	NA	NA	FADS1, FADS2	1.70E-02
KEGG One Carbon Pool By Folate	One carbon pool by folate	MTR, MTHFR	1.82E-02	NA	NA	DHFR, SHMT1, MTR, ATIC, MTHFR	3.01E-05	NA	NA
BioCarta RelA Pathway	Acetylation and Deacetylation of RelA in The Nucleus	TNFRSF1B, TNFRSF1A	1.72E-02	NA	NA	NA	NA	TNF, TNFRSF1A	1.16E-02
BioCarta TID Pathway	Chaperones modulate interferon Signaling Pathway	TNFRSF1B, TNFRSF1A	2.18E-02	NA	NA	NA	NA	TNF, TNFRSF1A	1.40E-02

Table 6 (continued).

KEGG Fc Gamma R Mediated Phagocytosis	Fc gamma R-mediated phagocytosis	NA	NA	NA	NA	AKT1, FCGR2A, DNM3, PRKCB, MYO10, WASL, PLA2G4A, GAB2, LIMK2, SYK, LYN	2.87E-06	NA	NA
BioCarta CardiacEGF Pathway	Role of EGF Receptor Transactivation by GPCRs in Cardiac Hypertrophy	NA	NA	NA	NA	PRKCB, EDN1, AGT, AGTR1, FOS	4.04E-05	NA	NA
BioCarta Free Pathway	Free Radical Induced Apoptosis	NA	NA	NA	NA	SOD1, XDH, TNF, IL8	6.40E-05	NA	NA
BioCarta ALK Pathway	ALK in cardiac myocytes	NA	NA	NA	NA	MEF2C, NOG, GSK3B, MYL2, TGFB1, BMP5	1.12E-04	NA	NA
KEGG TGF Beta Signaling Pathway	TGF-beta signaling pathway	NA	NA	NA	NA	NOG, TNF, INHBC, INHBA, TGFB1, ACVR2B, BMP5, BMP6	2.68E-04	NA	NA
BioCarta Gleevec Pathway	Inhibition of Cellular Proliferation by Gleevec	NA	NA	MAP3K1, AKT1, BAD, FOS	4.75E-04	NA	NA	NA	NA

Table 6 (continued).

KEGG Long Term Depression	Long-term depression	NA	NA	NA	NA	PRKCB, IGF1, CRHR1, PLA2G3, PLA2G4A, GNAS, LYN	4.61E-04	NA	NA
KEGG Glycerolipid Metabolism	Glycerolipid metabolism	NA	NA	NA	NA	DGKE, LPL, ALDH2, DGKB, LIPC, PNPLA3	4.63E-04	NA	NA
KEGG Glutathione Metabolism	Glutathione metabolism	NA	NA	NA	NA	GSTP1, GSTT1, RRM2B, RRM1, GCLC, GCLM	5.07E-04	NA	NA
KEGG Pyrimidine Metabolism	Pyrimidine metabolism	NA	NA	NA	NA	DCK, RRM2B, CDA, ITPA, RRM1, DPYD, UMPS, AK3	6.01E-04	NA	NA
BioCarta Fibrinolysis Pathway	Fibrinolysis Pathway	NA	NA	FGG, F13A1, FGA	1.16E-03	NA	NA	NA	NA
BioCarta AT1R Pathway	Angiotensin II mediated activation of JNK Pathway via Pyk2 dependent signaling	NA	NA	NA	NA	MEF2C, PRKCB, MAP3K1, AGT, AGTR1	8.62E-04	NA	NA
BioCarta BCR Pathway	BCR Signaling Pathway	NA	NA	NA	NA	PRKCB, SYK, MAP3K1, FOS, LYN	9.64E-04	NA	NA
KEGG Vibrio Cholerae Infection	Vibrio cholerae infection	ATP6V1H, ADCY3, ATP6V1C2, GNAS	2.25E-03	NA	NA	NA	NA	NA	NA

Table 6 (continued).

KEGG Thyroid Cancer	Thyroid cancer	CCND1, RXRG, PPARG	3.95E-03	NA	NA	NA	NA	NA	NA
KEGG Regulation Of Actin Cytoskeleton	Regulation of actin cytoskeleton	NA	NA	NA	NA	NA	NA	ARHGEF7, LIMK2, F2, ITGA2B, ITGA2	7.98E-03
BioCarta DC Pathway	Dendritic cells in regulating TH1 and TH2 Development	NA	NA	TLR4, TLR2, IL10	5.29E-03	NA	NA	NA	NA
BioCarta GCR Pathway	Corticosteroids and cardioprotection	NA	NA	GNAS, AKT1, ADRB2	5.29E-03	NA	NA	NA	NA
BioCarta HER2 Pathway	Role of ERBB2 in Signal Transduction and Oncology	NA	NA	IL6R, ESR1, IL6	5.29E-03	NA	NA	NA	NA
BioCarta IL6 Pathway	IL 6 signaling pathway	NA	NA	IL6R, FOS, IL6	5.29E-03	NA	NA	NA	NA
KEGG Beta Alanine Metabolism	beta-Alanine metabolism	NA	NA	ALDH2, ACADM, DPYD	5.29E-03	NA	NA	NA	NA
BioCarta Ptdins Pathway	Phosphoinositides and their downstream targets.	NA	NA	GSK3B, AKT1, BAD	5.91E-03	NA	NA	NA	NA
KEGG ErbB Signaling Pathway	ErbB signaling pathway	NA	NA	BAD, MAPK10, AKT1, GSK3B, NCK2	6.50E-03	NA	NA	NA	NA

Table 6 (continued).

KEGG Axon Guidance	Axon guidance	NA	NA	NFAT5, NCK2, SEMA5A, GSK3B, LIMK2, EPHA2	6.54E-03	NA	NA	NA	NA
KEGG Endometrial Cancer	Endometrial cancer	NA	NA	BAD, AKT1, GSK3B, CCND1	6.54E-03	NA	NA	NA	NA
KEGG Prostate Cancer	Prostate cancer	NA	NA	GSTP1, BAD, AKT1, GSK3B, CCND1	6.96E-03	NA	NA	NA	NA
KEGG Long Term Potentiation	Long-term potentiation	NA	NA	NA	NA	NA	NA	ADCY8, GRIN2A, CACNA1C	1.36E-02
KEGG Maturity Onset Diabetes Of The Young	Maturity onset diabetes of the young	NA	NA	SLC2A2, HNF4A, PAX4	7.09E-03	NA	NA	NA	NA
BioCarta EGFR SMRTE Pathway	Map Kinase Inactivation of SMRT Corepressor	MAP3K1, THRB	9.53E-03	NA	NA	NA	NA	NA	NA
BioCarta Lym Pathway	Adhesion and Diapedesis of Lymphocytes	IL1A, ICAM1	9.53E-03	NA	NA	NA	NA	NA	NA
KEGG Aldosterone Regulated Sodium Reabsorption	Aldosterone-regulated sodium reabsorption	KCNJ1, ATP1B1, NEDD4L	9.75E-03	NA	NA	NA	NA	NA	NA
BioCarta BAD Pathway	Regulation of BAD phosphorylation	NA	NA	KITLG, BAD, AKT1	7.87E-03	NA	NA	NA	NA

Table 6 (continued).

KEGG Insulin Signaling Pathway	Insulin signaling pathway	NA	NA	PPARGC1A, AKT1, BAD, MAPK10, GSK3B, PPP1R3B	8.30E-03	NA	NA	NA	NA
KEGG Glycerophospholipid Metabolism	Glycerophospholipid metabolism	NA	NA	NA	NA	NA	NA	PLA2G3, DGKB, PISD	1.62E-02
BioCarta AGPCR Pathway	Attenuation of GPCR Signaling	GNAS, GRK4	1.27E-02	NA	NA	NA	NA	NA	NA
KEGG Type Ii Diabetes Mellitus	Type II diabetes mellitus	SLC2A2, CACNA1C, KCNJ11	1.30E-02	NA	NA	NA	NA	NA	NA
BioCarta P53Hypoxia Pathway	Hypoxia and p53 in the Cardiovascular system	NA	NA	NA	NA	NA	NA	ABCB1, AKT1	1.79E-02
BioCarta NtHi Pathway	NFkB activation by Nontypeable Hemophilus influenzae	NA	NA	NA	NA	NA	NA	TNF, IL1B	1.92E-02
KEGG Progesterone Mediated Oocyte Maturation	Progesterone-mediated oocyte maturation	NA	NA	NA	NA	NA	NA	ADCY8, ADCY4, AKT1	1.96E-02
BioCarta Stress Pathway	TNF/Stress Related Signaling	NA	NA	NA	NA	NA	NA	TNF, TNFRSF1A	2.02E-02
BioCarta CTL Pathway	CTL mediated immune response against target cells	ICAM1, FAS	1.56E-02	NA	NA	NA	NA	NA	NA
BioCarta HSP27 Pathway	Stress Induction of HSP Regulation	IL1A, FAS	1.56E-02	NA	NA	NA	NA	NA	NA

Table 6 (continued).

BioCarta Pitx2 Pathway	Multi-step Regulation of Transcription by Pitx2	GSK3B, HDAC1	1.56E-02	NA	NA	NA	NA	NA	NA
KEGG Terpenoid Backbone Biosynthesis	Terpenoid backbone biosynthesis	FDPS, HMGCR	1.56E-02	NA	NA	NA	NA	NA	NA
BioCarta P53 Pathway	p53 Signaling Pathway	CCND1, CDK4	1.72E-02	NA	NA	NA	NA	NA	NA
KEGG Primary Bile Acid Biosynthesis	Primary bile acid biosynthesis	CYP7A1, CH25H	1.72E-02	NA	NA	NA	NA	NA	NA
BioCarta TNFR1 Pathway	TNFR1 Signaling Pathway	NA	NA	NA	NA	NA	NA	TNF, TNFRSF1A	2.54E-02
KEGG Hedgehog Signaling Pathway	Hedgehog signaling pathway	GSK3B, BMP2, BMP6	1.82E-02	NA	NA	NA	NA	NA	NA
BioCarta 41BB Pathway	The 4-1BB-dependent immune response	MAP3K5, MAP3K1	1.82E-02	NA	NA	NA	NA	NA	NA
BioCarta IL10 Pathway	IL-10 Anti-inflammatory Signaling Pathway	IL1A, IL10	1.82E-02	NA	NA	NA	NA	NA	NA
BioCarta Spry Pathway	Sprouty regulation of tyrosine kinase signals	SPRY1, CBL	2.00E-02	NA	NA	NA	NA	NA	NA
BioCarta TNFR2 Pathway	TNFR2 Signaling Pathway	MAP3K1, TNFRSF1B	2.00E-02	NA	NA	NA	NA	NA	NA
BioCarta Par1 Pathway	Thrombin signaling and protease-activated receptors	NA	NA	NA	NA	NA	NA	F2, ARHGEF7	3.81E-02

Table 6 (continued).

BioCarta Ceramide Pathway	Ceramide Signaling Pathway	MAP3K1, TNFRSF1A	2.77E-02	NA	NA	NA	NA	NA	NA

NA = No association / Not significant association

[1]Genes in the pathway that are associated with ancestry enriched SNPs

[2]FDR (q-value) computed from the gene set enrichment analysis (GSEA)

APPENDIX B. SUPPLEMENTARY TABLES FOR CHAPTER 5

Table 7. References and values for phenotypes with significant AMI values and population variance.

Phenotypes with significant AMI values						
Phenotype	**Population**	**Database[1]**	**PMID(s)[2]**	**AMI value[3]**	**q-value[4]**	**Wilcoxon P-value[5]**
BMI	Mexico	GIANT	20935630, 22982992, 23563607, 23754948, 25673413	0.27	2.42E-13	0
Schizophrenia	Mexico	GWAS Catalog	19571808, 19571811, 21682944, 21926974, 22037552, 22037555, 22688191, 22883433, 23142968, 23894747, 23974872, 24043878, 25056061, 26198764	0.26	6.99E-12	0
Body mass index	Mexico	GWAS Catalog	17434869, 18454148, 19079260, 19079261, 20935630, 22344219, 22344221, 22982992, 23563607, 23583978, 23669352, 24064335, 24348519, 24861553, 25673413, 25953783	0.24	4.08E-07	0
Height	Peru	GWAS Catalog	17767157, 18193045, 18391950, 18391951, 18391952, 18952825, 19343178, 19396169, 19570815, 19729412, 19893584, 20189936, 20397748, 20881960, 21998595, 22021425, 23456168, 23563607, 25282103, 25429064	0.13	3.62E-06	0

Table 7 (continued).

Height	Peru	GIANT	20881960, 22982992, 23563607, 23754948, 25282103	0.12	7.27E-05	0
BMI	Peru	GIANT	20935630, 22982992, 23563607, 23754948, 25673413	0.15	1.10E-04	0
Facial Development	Mexico	Custom	24651127	0.22	1.44E-03	1.46E-02
Height	Colombia	GIANT	20881960, 22982992, 23563607, 23754948, 25282103	0.11	1.60E-03	7.53E-243
Body mass index	Peru	GWAS Catalog	17434869, 18454148, 19079260, 19079261, 20935630, 22344219, 22344221, 22982992, 23563607, 23583978, 23669352, 24064335, 24348519, 24861553, 25673413, 25953783	0.15	2.11E-03	0
Schizophrenia	Peru	GWAS Catalog	19571808, 19571811, 21682944, 21926974, 22037552, 22037555, 22688191, 22883433, 23142968, 23894747, 23974872, 24043878, 25056061, 26198764	0.12	3.10E-03	7.28E-239
Educational Attainment	Mexico	Custom	27225129	0.22	4.32E-03	1.82E-10
Morning vs. evening chronotype	Mexico	GWAS Catalog	26955885, 26835600	0.22	1.92E-01	4.93E-29

Table 7 (continued).

Combined neurological disorders[6]	Mexico	GWAS Catalog	23453885	0.26	2.18E-01	0
Beard thickness	Mexico	GWAS Catalog	26926045	0.39	2.18E-01	0
Bipolar disorder	Mexico	GWAS Catalog	17554300, 17486107, 18711365, 19416921, 21926972, 21353194, 22205951, 22182935, 21254220, 24618891	0.23	2.18E-01	0
Childhood body mass index	Mexico	GWAS Catalog	26604143	0.24	2.18E-01	9.93E-20
Helix rolling	Mexico	GWAS Catalog	26105758	0.41	2.18E-01	8.17E-43
Lobe size	Mexico	GWAS Catalog	26105758	0.48	2.18E-01	0
Scalp hair shape	Mexico	GWAS Catalog	26926045	0.35	2.20E-01	0
Chronotype	Mexico	GWAS Catalog	26955885, 26835600	0.25	2.60E-01	0
Phenotypes with significant AMI population variance						
Phenotype		**Database[1]**	**PMID(s)[2]**	**AMI variance[7]**		***q*-value[8]**
Freckles		GWAS Catalog	23548203	1.41E-01		0
Sunburns		GWAS Catalog	17952075, 18488028	1.41E-01		0
Addiction		GWAS Catalog	23533358	7.23E-02		1.10E-103
HLA Class I		Custom	20356336	6.25E-02		2.20E-68
HLA Class II		Custom	20356336	6.25E-02		2.20E-68
Opioid sensitivity		GWAS Catalog	24143882	6.20E-02		6.53E-67
Lobe size		GWAS Catalog	26105758	5.85E-02		3.23E-56
Ear protrusion		GWAS Catalog	26105758	5.72E-02		1.67E-52

Table 7 (continued).

Eyebrow thickness	GWAS Catalog	26926045	5.72E-02	1.67E-52
Lobe attachment	GWAS Catalog	26105758	5.72E-02	1.67E-52
Monobrow thickness	GWAS Catalog	26926045	5.72E-02	1.67E-52
Height adjusted BMI	GWAS Catalog	25044758	5.36E-02	4.79E-43
Temperament (bipolar disorder)	GWAS Catalog	22365631	4.81E-02	6.37E-30
Drinking behavior	GWAS Catalog	21372407, 23364009	4.75E-02	8.38E-29
Helix rolling	GWAS Catalog	26105758	4.44E-02	1.04E-22
Ear morphology	GWAS Catalog	26105758	4.10E-02	6.61E-17
Body fat percentage	GWAS Catalog	26833246	8.83E-04	2.22E-16
Black vs. red hair color	GWAS Catalog	18483556	1.63E-03	2.96E-15
Vitiligo	GWAS Catalog	21326295, 22561518, 19890347, 20410501, 20526339, 22951725	1.94E-03	8.26E-15
Eye color	GWAS Catalog	20585627, 23118974, 23548203	2.27E-03	2.42E-14

[1] *Database source for the trait SNP-associations and gene sets. 'Custom' refers to SNP-associations mined from the literature.*
[2] *PubMed identifiers for the publications where the trait SNP-associations are reported.*
[3] *Value of the assortative mating index (AMI) test statistic for the phenotype gene set.*
[4] *False discovery rate q-value for the significance of the AMI value.*
[5] *P-value of the Wilcoxon rank sum test comparing the permuted AMI statistics to the observed AMI statistic*
[6] *Combined neurological disorders: Autism spectrum disorder, attention deficit-hyperactivity disorder, bipolar disorder, major depressive disorder, and schizophrenia (combined)*
[7] *Variance of the AMI test statistics for the phenotype gene set across all 4 Latin American populations.*
[8] *False discovery rate q-value for the variance of the AMI test statistics across the 4 Latin American populations.*

Table 8. Ancestry differences for phenotypes implicated in assortative mating (*i.e.* mate choice) in admixed Latin American populations.

Trait	Ancestry[1]			Variance	Source
	African	European	Native American		
Height (Male)	167 cm	175 cm	164 cm	34 cm	https://en.wikipedia.org/wiki/List_of_average_human_height_worldwide
Height (Female)	160 cm	162 cm	151 cm	38 cm	
Body Mass Index[2] (Male)	22.8	27.4	25.7	5.4	http://www.who.int/nmh/publications/ncd-status-report-2014/en/
Body Mass Index[2] (Female)	24.0	26.0	26.9	2.2	
Schizophrenia[3]	247	186	253	1374	http://www.who.int/healthinfo/global_burden_disease/2004_report_update/en/
Educational Attainment[4]	44%	67%	74%	246.3	https://www.education-inequalities.org/indicators/comp_upsec_v2

[1] *Ancestry-specific measures are based on values from Nigeria (African), Spain (European), and Peru (Native American). Nigeria and Spain are chosen, as they are known to provide the highest African and European ancestry components to admixed Latin American countries. Peru is chosen as it is known to have the highest Native American ancestry proportion in Latin America (ref http://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1005602).*

[2] *Body mass index is measured as a person's mass (weight) divided by the square of their height.*

[3] *Schizophrenia is measured as the age-standardized disability-adjusted life years (DALY) rates per 100,000 inhabitants (recorded in 2004).*

[4] *Educational attainment is the percentage of people aged 3-5 years above upper secondary school graduation age who have completed upper secondary school.*

PUBLICATIONS

1. **Norris, E.T.**, Rishishwar, L., Chande, A.T., Conley, A.B., Ye, K., Valderrama-Aguirre, A., Jordan, I.K. Admixture-enabled selection for rapid adaptive evolution in the Americas. *In review*
2. **Norris, E.T.**, Rishishwar, L., Jordan, I.K. Rapid, Adaptive Human Evolution Facilitated by Admixture in the Americas. In: Human Migration: Biocultural Perspectives. Oxford University Press, USA. *In press*.
3. **Norris, E.T.**, Rishishwar, L., Wang, L., Conley, A.B., Chande, A.T., Dabrowski, A.M., Valderrama-Aguirre, A., Jordan, I.K. (2019). Assortative mating on ancestry-variant traits in admixed Latin American populations. Front Genet. doi: 10.3389/fgene.2019.00359.
4. **Norris, E.T.**, Wang, L., Conley, A.B., Rishishwar, L., Mariño-Ramírez, L., Valderrama-Aguirre, A., Jordan I.K. (2018). Genetic ancestry, admixture and health determinants in Latin America. BMC Genomics. 19(Suppl 8):861.

5. Nagar, S.D., Moreno, A.M., **Norris, E.T.**, Rishishwar, L., Conley, A.B., O'Neal, K.L., Vélez-Gómez, S., Montes-Rodríguez, C., Jaraba-Álvarez, W.V., Torres, I., Medina-Rivas, M.A., Valderrama-Aguirre, A., Jordan, I.K., Gallo, J.E. (2019). Population Pharmacogenomics for Precision Public Health in Colombia. Front Genet. doi: 10.3389/fgene.2019.00241.
6. Chande, A.T., Wang, L., Rishishwar, L., Conley, A.B., **Norris, E.T.**, Valderrama-Aguirre, A., Jordan, I.K. (2018). GlobAl Distribution of GEnetic Traits (GADGET)

web server: polygenic trait scores worldwide. Nucleic Acids Res. 46(W1):W121-W126.

7. Chande, A.T., Rowell, J., Rishishwar, L., Conley, A.B., **Norris, E.T.**, Valderrama-Aguirre, A., Medina-Rivas, M.A., Jordan, I.K. (2017). Influence of genetic ancestry and socioeconomic status on type 2 diabetes in the diverse Colombian populations of Chocó and Antioquia. Sci Rep. doi: 10.1038/s41598-017-17380-4.
8. Conley, A.B., Rishishwar, L., **Norris, E.T.**, Valderrama-Aguirre, A., Mariño-Ramírez, L., Medina-Rivas, M.A., Jordan, I.K. (2017). A Comparative Analysis of Genetic Ancestry and Admixture in the Colombian Populations of Chocó and Medellín. G3 (Bethesda). doi: 10.1534/g3.117.1118.
9. Wang, L., **Norris, E.T.**, Jordan, I.K. (2017). Human Retrotransposon Insertion Polymorphisms Are Associated with Health and Disease via Gene Regulatory Phenotypes. Front Microbiol. doi: 10.3389/fmicb.2017.01418.
10. Medina-Rivas, M.A., **Norris, E.T.**, Rishishwar, L., Conley, A.B., Medrano-Trochez, C., Valderrama-Aguirre, A., Vannberg, F.O., Mariño-Ramírez, L., Jordan, I.K. (2016). Chocó, Colombia: a hotspot of human biodiversity. Rev Biodivers Neotrop. 6(1):45-54.

REFERENCES

1. Schraiber, J.G. and J.M. Akey, *Methods and models for unravelling human evolutionary history.* Nat Rev Genet, 2015. **16**(12): p. 727-40.

2. Veeramah, K.R. and M.F. Hammer, *The impact of whole-genome sequencing on the reconstruction of human population history.* Nat Rev Genet, 2014. **15**(3): p. 149-62.

3. Crosby, A.W., *The Columbian exchange: biological and cultural consequences of 1492*. Vol. 2. 2003, Westport: Greenwood Publishing Group.

4. Mann, C.C., *1493: Uncovering the new world Columbus created*. 2011, New York: Alfred a Knopf Incorporated.

5. Chande, A.T., et al., *Influence of genetic ancestry and socioeconomic status on type 2 diabetes in the diverse Colombian populations of Choco and Antioquia.* Sci Rep, 2017. **7**(1): p. 17127.

6. Nagar, S.D., et al., *Population Pharmacogenomics for Precision Public Health in Colombia.* Front Genet, 2019. **10**: p. 241.

7. Need, A.C. and D.B. Goldstein, *Next generation disparities in human genomics: concerns and remedies.* Trends Genet, 2009. **25**(11): p. 489-94.

8. Popejoy, A.B. and S.M. Fullerton, *Genomics is failing on diversity.* Nature, 2016. **538**(7624): p. 161-164.

9. Currat, M., et al., *Molecular analysis of the beta-globin gene cluster in the Niokholo Mandenka population reveals a recent origin of the beta(S) Senegal mutation.* Am J Hum Genet, 2002. **70**(1): p. 207-23.

10. Ohashi, J., et al., *Extended linkage disequilibrium surrounding the hemoglobin E variant due to malarial selection.* Am J Hum Genet, 2004. **74**(6): p. 1198-208.

11. Enattah, N.S., et al., *Identification of a variant associated with adult-type hypolactasia.* Nat Genet, 2002. **30**(2): p. 233-7.

12. Bersaglieri, T., et al., *Genetic signatures of strong recent positive selection at the lactase gene.* Am J Hum Genet, 2004. **74**(6): p. 1111-20.

13. Tishkoff, S.A., et al., *Convergent adaptation of human lactase persistence in Africa and Europe.* Nat Genet, 2007. **39**(1): p. 31-40.

14. Huerta-Sanchez, E., et al., *Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA.* Nature, 2014. **512**(7513): p. 194-7.

15. Jeong, C., et al., *Admixture facilitates genetic adaptations to high altitude in Tibet.* Nat Commun, 2014. **5**: p. 3281.

16. Bigham, A., et al., *Identifying signatures of natural selection in Tibetan and Andean populations using dense genome scan data.* PLoS Genet, 2010. **6**(9): p. e1001116.

17. Jordan, I.K., *The Columbian Exchange as a source of adaptive introgression in human populations.* Biol Direct, 2016. **11**(1): p. 17.

18. Buss, D.M., *Human mate selection: Opposites are sometimes said to attract, but in fact we are likely to marry someone who is similar to us in almost every variable.* American scientist, 1985. **73**(1): p. 47-51.

19. Robinson, M.R., et al., *Genetic evidence of assortative mating in humans.* Nature Human Behaviour, 2017. **1**: p. 0016.

20. Vandenburg, S.G., *Assortative mating, or who marries whom?* Behav Genet, 1972. **2**(2): p. 127-57.

21. Fleagle, J.G., et al., *Out of Africa I: the first hominin colonization of Eurasia*. 2010: Springer Science & Business Media.

22. Zhu, Z., et al., *Hominin occupation of the Chinese Loess Plateau since about 2.1 million years ago.* Nature, 2018. **559**(7715): p. 608-612.

23. Schlebusch, C.M., et al., *Southern African ancient genomes estimate modern human divergence to 350,000 to 260,000 years ago.* Science, 2017. **358**(6363): p. 652-655.

24. Henn, B.M., L.L. Cavalli-Sforza, and M.W. Feldman, *The great human expansion.* Proc Natl Acad Sci U S A, 2012. **109**(44): p. 17758-64.

25. Nielsen, R., et al., *Tracing the peopling of the world through genomics.* Nature, 2017. **541**(7637): p. 302-310.

26. Reich, D., *Who we are and how we got here: Ancient DNA and the new science of the human past*. 2018: Oxford University Press.

27. Hellenthal, G., et al., *A genetic atlas of human admixture history.* Science, 2014. **343**(6172): p. 747-751.

28. Wright, S.W. *The roles of mutation, inbreeding, crossbreeding and selection in evolution*. in *Proceedings of the 6th International Congress of Genetics*. 1932.

29. Cochran, G. and H. Harpending, *The 10,000 year explosion: How civilization accelerated human evolution*. 2009: Basic Books.

30. Hawks, J., et al., *Recent acceleration of human adaptive evolution.* Proc Natl Acad Sci U S A, 2007. **104**(52): p. 20753-8.

31. Vitti, J.J., S.R. Grossman, and P.C. Sabeti, *Detecting natural selection in genomic data.* Annu Rev Genet, 2013. **47**: p. 97-120.

32. Sabeti, P.C., et al., *Positive natural selection in the human lineage.* Science, 2006. **312**(5780): p. 1614-20.

33. Oleksyk, T.K., M.W. Smith, and S.J. O'Brien, *Genome-wide scans for footprints of natural selection.* Philos Trans R Soc Lond B Biol Sci, 2010. **365**(1537): p. 185-205.

34. Fan, S., et al., *Going global by adapting local: A review of recent human adaptation.* Science, 2016. **354**(6308): p. 54-59.

35. Akey, J.M., et al., *Population history and natural selection shape patterns of genetic variation in 132 genes.* PLoS Biol, 2004. **2**(10): p. e286.

36. Helgason, A., et al., *Refining the impact of TCF7L2 gene variants on type 2 diabetes and adaptive evolution.* Nat Genet, 2007. **39**(2): p. 218-25.

37. Field, Y., et al., *Detection of human adaptation during the past 2000 years.* Science, 2016. **354**(6313): p. 760-764.

38. Gibbons, A., *European skin turned pale only recently, gene suggests.* Science, 2007. **316**(5823): p. 364.

39. Shriner, D. and C.N. Rotimi, *Whole-Genome-Sequence-Based Haplotypes Reveal Single Origin of the Sickle Allele during the Holocene Wet Phase.* Am J Hum Genet, 2018. **102**(4): p. 547-556.

40. Racimo, F., et al., *Evidence for archaic adaptive introgression in humans.* Nat Rev Genet, 2015. **16**(6): p. 359-71.

41. Abi-Rached, L., et al., *The shaping of modern human immune systems by multiregional admixture with archaic humans.* Science, 2011. **334**(6052): p. 89-94.

42. Mendez, F.L., J.C. Watkins, and M.F. Hammer, *A haplotype at STAT2 Introgressed from neanderthals and serves as a candidate of positive selection in Papua New Guinea.* Am J Hum Genet, 2012. **91**(2): p. 265-74.

43. Sankararaman, S., et al., *The genomic landscape of Neanderthal ancestry in present-day humans.* Nature, 2014. **507**(7492): p. 354-7.

44. Vernot, B. and J.M. Akey, *Resurrecting surviving Neandertal lineages from modern human genomes.* Science, 2014. **343**(6174): p. 1017-21.

45. Dannemann, M., A.M. Andres, and J. Kelso, *Introgression of Neandertal- and Denisovan-like Haplotypes Contributes to Adaptive Variation in Human Toll-like Receptors.* Am J Hum Genet, 2016. **98**(1): p. 22-33.

46. Beall, C.M., et al., *Natural selection on EPAS1 (HIF2alpha) associated with low hemoglobin concentration in Tibetan highlanders.* Proc Natl Acad Sci U S A, 2010. **107**(25): p. 11459-64.

47. Simonson, T.S., et al., *Genetic evidence for high-altitude adaptation in Tibet.* Science, 2010. **329**(5987): p. 72-5.

48. Yi, X., et al., *Sequencing of 50 human exomes reveals adaptation to high altitude.* Science, 2010. **329**(5987): p. 75-8.

49. Patin, E., et al., *Dispersals and genetic adaptation of Bantu-speaking populations in Africa and North America.* Science, 2017. **356**(6337): p. 543-546.

50. Norris, E.T., et al., *Genetic ancestry, admixture and health determinants in Latin America.* BMC Genomics, 2018. **19**(Suppl 8): p. 861.

51. Maples, B.K., et al., *RFMix: a discriminative modeling approach for rapid and robust local-ancestry inference.* Am J Hum Genet, 2013. **93**(2): p. 278-88.

52. Geza, E., et al., *A comprehensive survey of models for dissecting local ancestry deconvolution in human genome.* Brief Bioinform, 2018.

53. Tang, H., et al., *Recent genetic selection in the ancestral admixture of Puerto Ricans.* Am J Hum Genet, 2007. **81**(3): p. 626-33.

54. Gilad, Y., et al., *Natural selection on the olfactory receptor gene family in humans and chimpanzees.* Am J Hum Genet, 2003. **73**(3): p. 489-501.

55. Voight, B.F., et al., *A map of recent positive selection in the human genome.* PLoS Biol, 2006. **4**(3): p. e72.

56. Rishishwar, L., et al., *Ancestry, admixture and fitness in Colombian genomes.* Sci Rep, 2015. **5**: p. 12376.

57. Guan, Y., *Detecting structure of haplotypes and local ancestry.* Genetics, 2014. **196**(3): p. 625-42.

58. Zhou, Q., L. Zhao, and Y. Guan, *Strong Selection at MHC in Mexicans since Admixture.* PLoS Genet, 2016. **12**(2): p. e1005847.

59. Jin, W., et al., *Genome-wide detection of natural selection in African Americans pre- and post-admixture.* Genome Res, 2012. **22**(3): p. 519-27.

60. Bhatia, G., et al., *Genome-wide scan of 29,141 African Americans finds no evidence of directional selection since admixture.* Am J Hum Genet, 2014. **95**(4): p. 437-44.

61. Gould, S.J., *The spice of life.* Leader to Leader, 2000. **15**: p. 14-19.

62. Mayr, E., *Animal species and evolution*. 1963: Belknap Press of Harvard University Press.

63. Rosenberg, N.A., et al., *Genetic structure of human populations.* Science, 2002. **298**(5602): p. 2381-5.

64. Li, J.Z., et al., *Worldwide human relationships inferred from genome-wide patterns of variation.* Science, 2008. **319**(5866): p. 1100-4.

65. Ruiz-Linares, A., et al., *Admixture in Latin America: geographic structure, phenotypic diversity and self-perception of ancestry based on 7,342 individuals.* PLoS Genet, 2014. **10**(9): p. e1004572.

66. Wang, S., et al., *Geographic patterns of genome admixture in Latin American Mestizos.* PLoS Genet, 2008. **4**(3): p. e1000037.

67. Bryc, K., et al., *Colloquium paper: genome-wide patterns of population structure and admixture among Hispanic/Latino populations.* Proc Natl Acad Sci U S A, 2010. **107 Suppl 2**: p. 8954-61.

68. Bryc, K., et al., *Genome-wide patterns of population structure and admixture in West Africans and African Americans.* Proc Natl Acad Sci U S A, 2010. **107**(2): p. 786-91.

69. Basu, A., et al., *Genome-wide distribution of ancestry in Mexican Americans.* Hum Genet, 2008. **124**(3): p. 207-14.

70. Genomes Project, C., et al., *A global reference for human genetic variation.* Nature, 2015. **526**(7571): p. 68-74.

71. Chang, C.C., et al., *Second-generation PLINK: rising to the challenge of larger and richer datasets.* Gigascience, 2015. **4**: p. 7.

72. Delaneau, O., J.F. Zagury, and J. Marchini, *Improved whole-chromosome phasing for disease and population genetic studies.* Nat Methods, 2013. **10**(1): p. 5-6.

73. Alexander, D.H., J. Novembre, and K. Lange, *Fast model-based estimation of ancestry in unrelated individuals.* Genome Res, 2009. **19**(9): p. 1655-64.

74. O'Leary, N.A., et al., *Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation.* Nucleic Acids Res, 2016. **44**(D1): p. D733-45.

75. Subramanian, A., et al., *Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles.* Proc Natl Acad Sci U S A, 2005. **102**(43): p. 15545-50.

76. Ritchie, M.E., et al., *limma powers differential expression analyses for RNA-sequencing and microarray studies.* Nucleic Acids Res, 2015. **43**(7): p. e47.

77. Lappalainen, T., et al., *Transcriptome and genome sequencing uncovers functional variation in humans.* Nature, 2013. **501**(7468): p. 506-11.

78. Flicek, P., et al., *Ensembl 2013.* Nucleic Acids Res, 2013. **41**(Database issue): p. D48-55.

79. t Hoen, P.A., et al., *Reproducibility of high-throughput mRNA and small RNA sequencing across laboratories.* Nat Biotechnol, 2013. **31**(11): p. 1015-22.

80. Stegle, O., et al., *Using probabilistic estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses.* Nat Protoc, 2012. **7**(3): p. 500-7.

81. Shabalin, A.A., *Matrix eQTL: ultra fast eQTL analysis via large matrix operations.* Bioinformatics, 2012. **28**(10): p. 1353-8.

82. Viechtbauer, W., *Conducting meta-analyses in R with the metafor package.* Journal of Statistical Software, 2010. **36**(3): p. 1-48.

83. Gilad, Y., S.A. Rifkin, and J.K. Pritchard, *Revealing the architecture of gene regulation: the promise of eQTL studies.* Trends Genet, 2008. **24**(8): p. 408-15.

84. Gibson, G., J.E. Powell, and U.M. Marigorta, *Expression quantitative trait locus analysis for translational medicine.* Genome Med, 2015. **7**(1): p. 60.

85. Conley, A.B., et al., *A Comparative Analysis of Genetic Ancestry and Admixture in the Colombian Populations of Choco and Medellin.* G3 (Bethesda), 2017. **7**(10): p. 3435-3447.

86. Medina-Rivas, M.A., et al., *Choco, Colombia: a hotspot of human biodiversity.* Rev Biodivers Neotrop, 2016. **6**(1): p. 45-54.

87. Lopez, M., et al., *Genomic Evidence for Local Adaptation of Hunter-Gatherers to the African Rainforest.* Curr Biol, 2019. **29**(17): p. 2926-2935 e4.

88. Moreno-Estrada, A., et al., *Reconstructing the population genetic history of the Caribbean.* PLoS Genet, 2013. **9**(11): p. e1003925.

89. Homburger, J.R., et al., *Genomic Insights into the Ancestry and Demographic History of South America.* PLoS Genet, 2015. **11**(12): p. e1005602.

90. Boyko, A.R., et al., *Assessing the evolutionary impact of amino acid mutations in the human genome.* PLoS Genet, 2008. **4**(5): p. e1000083.

91. Fay, J.C., G.J. Wyckoff, and C.I. Wu, *Positive and negative selection on the human genome.* Genetics, 2001. **158**(3): p. 1227-34.

92. Drake, J.W., et al., *Rates of spontaneous mutation.* Genetics, 1998. **148**(4): p. 1667-86.

93. Messer, P.W. and D.A. Petrov, *Population genomics of rapid adaptation by soft selective sweeps.* Trends Ecol Evol, 2013. **28**(11): p. 659-69.

94. Peter, B.M., E. Huerta-Sanchez, and R. Nielsen, *Distinguishing between selective sweeps from standing variation and from a de novo mutation.* PLoS Genet, 2012. **8**(10): p. e1003011.

95. Racimo, F., J.J. Berg, and J.K. Pickrell, *Detecting Polygenic Adaptation in Admixture Graphs.* Genetics, 2018. **208**(4): p. 1565-1584.

96. Berg, J.J. and G. Coop, *A population genetic signal of polygenic adaptation.* PLoS Genet, 2014. **10**(8): p. e1004412.

97. Turchin, M.C., et al., *Evidence of widespread selection on standing variation in Europe at height-associated SNPs.* Nat Genet, 2012. **44**(9): p. 1015-9.

98. Beiter, E.R., et al., *Polygenic selection underlies evolution of human brain structure and behavioral traits.* bioRxiv, 2017: p. 164707.

99. Sudmant, P.H., et al., *An integrated map of structural variation in 2,504 human genomes.* Nature, 2015. **526**(7571): p. 75-81.

100. Reich, D., et al., *Reconstructing Native American population history.* Nature, 2012. **488**(7411): p. 370-4.

101. Jordan, I.K., L. Rishishwar, and A.B. Conley, *Native American admixture recapitulates population-specific migration and settlement of the continental United States.* PLoS Genet, 2019. **15**(9): p. e1008225.

102. Chande, A.T., Conley, A.B., Jordan, I.K., Rishishwar, L, *Tagore: a utility for illustrating human chromosomes*. In preparation.

103. Haeussler, M., et al., *The UCSC Genome Browser database: 2019 update.* Nucleic Acids Res, 2019. **47**(D1): p. D853-D858.

104. Benjamini, Y. and Y. Hochberg, *Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing.* Journal of the Royal Statistical Society. Series B (Methodological), 1995. **57**(1): p. 289-300.

105. Szpiech, Z.A. and R.D. Hernandez, *selscan: an efficient multithreaded program to perform EHH-based scans for positive selection.* Mol Biol Evol, 2014. **31**(10): p. 2824-7.

106. Hartl, D. and A. Clark, *Principles of population genetics*. 1989, Sunderland, Massachusetts: Sinauer Associates.

107. MacArthur, J., et al., *The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog).* Nucleic Acids Res, 2017. **45**(D1): p. D896-D901.

108. Li, S., et al., *Molecular signatures of antibody responses derived from a systems biology study of five human vaccines.* Nat Immunol, 2014. **15**(2): p. 195-204.

109. Breuer, K., et al., *InnateDB: systems biology of innate immunity and beyond--recent updates and continuing curation.* Nucleic Acids Res, 2013. **41**(Database issue): p. D1228-33.

110. Preininger, M., et al., *Blood-informative transcripts define nine common axes of peripheral blood gene expression.* PLoS Genet, 2013. **9**(3): p. e1003362.

111. Deschamps, M., et al., *Genomic Signatures of Selective Pressures and Introgression from Archaic Hominins at Human Innate Immunity Genes.* Am J Hum Genet, 2016. **98**(1): p. 5-21.

112. Consortium, G. *GIANT: Genetic Investigation of ANthropometric Traits*. November 1, 2018]; Available from: http://portals.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium.

113. Cavalli-Sforza, L.L., P. Menozzi, and A. Piazza, *The history and geography of human genes*. 1994, Princeton: Princeton University Press.

114. Hellenthal, G., et al., *A genetic atlas of human admixture history.* Science, 2014. **343**(6172): p. 747-51.

115. Montinaro, F., et al., *Unravelling the hidden ancestry of American admixed populations.* Nat Commun, 2015. **6**: p. 6596.

116. Allison, D.B., et al., *Assortative mating for relative weight: genetic implications.* Behav Genet, 1996. **26**(2): p. 103-11.

117. Silventoinen, K., et al., *Assortative mating by body height and BMI: Finnish twins and their spouses.* Am J Hum Biol, 2003. **15**(5): p. 620-7.

118. Stulp, G., et al., *Assortative mating for human height: A meta-analysis.* Am J Hum Biol, 2017. **29**(1).

119. Merikangas, K.R., *Assortative mating for psychiatric disorders and psychological traits.* Arch Gen Psychiatry, 1982. **39**(10): p. 1173-80.

120. Mare, R.D., *Five decades of educational assortative mating.* American sociological review, 1991: p. 15-32.

121. Domingue, B.W., et al., *Genetic and educational assortative mating among US adults.* Proc Natl Acad Sci U S A, 2014. **111**(22): p. 7996-8000.

122. Hur, Y.M., *Assortative mating for personality traits, educational level, religious affiliation, height, weight, and body mass index in parents of Korean twin sample.* Twin Res, 2003. **6**(6): p. 467-70.

123. Salces, I., E. Rebato, and C. Susanne, *Evidence of phenotypic and social assortative mating for anthropometric and physiological traits in couples from the Basque country (Spain).* J Biosoc Sci, 2004. **36**(2): p. 235-50.

124. Zou, J.Y., et al., *Genetic and socioeconomic study of mate choice in Latinos reveals novel assortment patterns.* Proc Natl Acad Sci U S A, 2015. **112**(44): p. 13621-6.

125. Kandler, C., W. Bleidorn, and R. Riemann, *Left or right? Sources of political orientation: the roles of genetic factors, cultural transmission, assortative mating, and personality.* J Pers Soc Psychol, 2012. **102**(3): p. 633-45.

126. Kalmijn, M., *Assortative mating by cultural and economic occupational status.* American Journal of Sociology, 1994. **100**(2): p. 422-452.

127. Chaix, R., C. Cao, and P. Donnelly, *Is mate choice in humans MHC-dependent?* PLoS Genet, 2008. **4**(9): p. e1000184.

128. Wedekind, C., et al., *MHC-dependent mate preferences in humans.* Proc Biol Sci, 1995. **260**(1359): p. 245-9.

129. Lykken, D.T. and A. Tellegen, *Is human mating adventitious or the result of lawful choice? A twin study of mate selection.* J Pers Soc Psychol, 1993. **65**(1): p. 56-68.

130. Rushton, J.P. and T.A. Bons, *Mate choice and friendship in twins: evidence for genetic similarity.* Psychol Sci, 2005. **16**(7): p. 555-9.

131. Zietsch, B.P., et al., *Variation in human mate choice: simultaneously investigating heritability, parental influence, sexual imprinting, and assortative mating.* Am Nat, 2011. **177**(5): p. 605-16.

132. Verweij, K.J., A.V. Burri, and B.P. Zietsch, *Evidence for genetic variation in human mate preferences for sexually dimorphic physical traits.* PLoS One, 2012. **7**(11): p. e49294.

133. Li, X., et al., *Height associated variants demonstrate assortative mating in human populations.* Sci Rep, 2017. **7**(1): p. 15689.

134. Sebro, R., et al., *Testing for non-random mating: evidence for ancestry-related assortative mating in the Framingham heart study.* Genet Epidemiol, 2010. **34**(7): p. 674-9.

135. Zaitlen, N., et al., *The Effects of Migration and Assortative Mating on Admixture Linkage Disequilibrium.* Genetics, 2017. **205**(1): p. 375-383.

136. Risch, N., et al., *Ancestry-related assortative mating in Latino populations.* Genome Biol, 2009. **10**(11): p. R132.

137. Hancock, A.M., et al., *Adaptations to climate-mediated selective pressures in humans.* PLoS Genet, 2011. **7**(4): p. e1001375.

138. Geary, D.C., *Evolution and proximate expression of human paternal investment.* Psychol Bull, 2000. **126**(1): p. 55-77.

139. Buss, D.M., *Conflict between the sexes: strategic interference and the evocation of anger and upset.* J Pers Soc Psychol, 1989. **56**(5): p. 735-47.

140. Procidano, M.E. and L.H. Rogler, *Homogamous assortative mating among Puerto Rican families: intergenerational processes and the migration experience.* Behav Genet, 1989. **19**(3): p. 343-54.

141. Trachtenberg, A., et al., *Canonical correlation analysis of assortative mating in two groups of Brazilians.* J Biosoc Sci, 1985. **17**(4): p. 389-403.

142. Malina, R.M., et al., *Assortative mating for phenotypic characteristics in a Zapotec community in Oaxaca, Mexico.* J Biosoc Sci, 1983. **15**(3): p. 273-80.

143. Frisancho, A.R., R. Wainwright, and A. Way, *Heritability and components of phenotypic expression in skin reflectance of Mestizos from the Peruvian lowlands.* Am J Phys Anthropol, 1981. **55**(2): p. 203-8.

144. Saha, S., et al., *A systematic review of the prevalence of schizophrenia.* PLoS Med, 2005. **2**(5): p. e141.

145. Davies, G., et al., *Genome-wide association study of cognitive functions and educational attainment in UK Biobank (N=112 151).* Mol Psychiatry, 2016. **21**(6): p. 758-67.

146. Okbay, A., et al., *Genome-wide association study identifies 74 loci associated with educational attainment.* Nature, 2016. **533**(7604): p. 539-42.

147. Rietveld, C.A., et al., *Common genetic variants associated with cognitive performance identified using the proxy-phenotype method.* Proc Natl Acad Sci U S A, 2014. **111**(38): p. 13790-4.

148. Rietveld, C.A., et al., *GWAS of 126,559 individuals identifies genetic variants associated with educational attainment.* Science, 2013. **340**(6139): p. 1467-71.

149. Jacob, S., et al., *Paternally inherited HLA alleles are associated with women's choice of male odor.* Nat Genet, 2002. **30**(2): p. 175-9.

150. Sebro, R., et al., *Structured mating: Patterns and implications.* PLoS Genet, 2017. **13**(4): p. e1006655.

151. Abdellaoui, A., K.J. Verweij, and B.P. Zietsch, *No evidence for genetic assortative mating beyond that due to population stratification.* Proc Natl Acad Sci U S A, 2014. **111**(40): p. E4137.

152. Gravel, S., *Population genetics models of local ancestry.* Genetics, 2012. **191**(2): p. 607-19.

153. Gravel, S., et al., *Reconstructing Native American migrations from whole-genome and whole-exome data.* PLoS Genet, 2013. **9**(12): p. e1004023.

154. Lachance, J., *A fundamental relationship between genotype frequencies and fitnesses.* Genetics, 2008. **180**(2): p. 1087-93.